FINAL REPORT

DOE EMSP
AWARD NUMBER: DE-FG07-02ER63500

AWARDEE NAME: University of Idaho, Moscow

PROJECT OR AWARDEE ACCOUNT NO.: EMSP 87003

PROJECT TITLE: Coupling of Realistic Real Estimates with Genomics for Assessing Contaminant Attenuation and Long-Term Plume Containment

December 13, 2006

Prepared by: Ronald L. Crawford, PhD
Director
University of Idaho
Environmental Biotechnology Institute
FRC 204
Moscow, ID 83844-1052
Phone: 208-885-6580
Fax: 208-885-5741
E-mail: crawford@uidaho.edu

NARRATIVE

ACCOMPLISHMENTS

Overview

This report is a summary final report of the research undertaken at the University of Idaho (UI) over the past four years under DOE EMSP award DE-FG07-02ER63500. This work was performed collaboratively with investigators at the Idaho National Laboratory (INL) and Northwind Inc., who will provide separate summaries of their portions of the interdisciplinary project. The research centered on studies of the subsurface environment of the Snake River Plain Aquifer (SRPA) of the Test Area North (TAN) site at the INL in southeastern Idaho. Through this interdisciplinary effort, UI, INL, and Northwind, Inc. investigated coupled biogeochemical processes implicit in conceptual and mathematical models that describe trichloroethylene (TCE) co-metabolism in complex subsurface strata. Our team accomplished the following:

- Investigated the microbial ecology of the SRPA related to known TCE co-metabolizers.
- Obtained evidence of biogenic methane in the SRPA that can sustain methanotrophic cometabolism of TCE.
- Developed conceptual and mathematical models for natural attenuation at TAN and other DOE sites.
- Used specific enzyme activities to detect TCE co-metabolic potential in SRPA microbes.
- Acquired stable carbon isotope fractionation data for contaminants that demonstrate enhanced bioremediation of TCE.
- Gained extensive field experience in subsurface microbiology, bioremediation, and design and operation of *in situ* bioreactors.
- Two students received graduate degrees through their participation on the project.
- One postdoctoral scientist received training through participation on the project.

The following publications and presentations co-authored by members of the University of Idaho group were partially or completely made possible through the funding provided under award DE-FG07-02ER63500.

Publications

1. Sebat, J. L., F. S. Colwell, and R. L. Crawford. 2003. Metagenomic Profiling: Microarray analysis of an environmental genomic library: Comparative genomic hybridization. ***Appl. Environ. Microbiol.*** 69:4927-4934.
2. Erwin, D. P., I. K. Erickson, M. E. Delwiche, F. S. Colwell, J. L. Strap, and R. L. Crawford. 2005. Diversity of oxygenase genes from methane- and ammonia-oxidizing bacteria in the Eastern Snake River Plain aquifer. ***Appl. Environ. Microbiol.*** 71:2016-2025.
3. Strap, J. L., F. S. Colwell and R. L. Crawford. Eukaryotic diversity in planktonic and biofilm populations of the Snake River Plain Aquifer. ***Microbial Ecology.*** In review.

Presentations

1. Strap, J. L., F. S. Colwell, and R. L. Crawford. Comparison of planktonic and biofilm microbial diversity found within the Snake River Plain Aquifer. Joint Intl. Symp. for Subsurface Microbiology and Environmental Biogeochemistry, Jackson Hole, WY; Aug. 14-19, 2005.
2. Crawford, R. Genomic analysis of methane metabolizing bacterial communities in groundwater. INRA Subsurface Science Symposium, Boise, ID; October 13-16, 2002.
3. Sebat, J. Classification of DNA from a metagenomic library by microarray comparative genomic hybridization. INRA Subsurface Science Symposium, Boise, ID; October 13-16, 2002.
4. Crawford, R. L., J. L. Sebat, and F. S. Colwell. Classification of DNA from a multigenomic library using microarray comparative genomic hybridization. Am. Soc. Microbiology Annual meeting, Washington, DC; May 18-22, 2003.
5. Crawford, R. L., J. L. Sebat, and F. S. Colwell. Classification of DNA from a multigenomic library using microarray comparative genomic hybridization. Invited speaker, Northwest Branch ASM meeting, Vancouver, BC, Canada; August 8-10, 2003.
6. Crawford, R. L., J. L. Sebat, and F. S. Colwell. Genomic analysis of a microbial community by microarray comparative genomic hybridization. Invited speaker, Northwest Gene Expression Conference, Seattle, WA; August 27-29, 2003.
7. Crawford, R. L., J. L. Sebat, and F. S. Colwell. Genomic analysis of a microbial community by microarray comparative genomic hybridization. INRA Subsurface Science Symposium, Salt Lake City, UT; October 6-8, 2003.
8. Erwin, D. P., and R. L. Crawford. Genomic analysis of subsurface microbial communities of the SRPA. INRA Subsurface Science Symposium, Salt Lake City, UT; October 6-8, 2003.
9. Strap, J. L., A. L. Torguson, F. S. Colwell, and R. L. Crawford. Characterization of the microbial diversity associated with the planktonic and biofilm populations in a deep-water well from the Snake River Plain Aquifer. Am Soc. Microbiology Annual meeting, New Orleans, LA; May 23-27, 2004.
10. Boyd, S. S., F. S. Colwell, D. W. Reed, M. E. Delwiche, and R. Crawford. Determination of methanogen biomass in hydrate-bearing sediments by quantitative PCR. Am Soc. Microbiology Annual meeting, New Orleans, LA; May 23-27, 2004.
11. Strap, J. L., F. S. Colwell, and R. L. Crawford. Microbial diversity of subsurface communities of the SRP aquifer. INRA Subsurface Science Symposium, Spokane, WA; Sept. 20-22, 2004.
12. Strap, J. L., F. S. Colwell, and R. L. Crawford. Characterization of archaeal diversity associated with planktonic and biofilm subsurface communities from the Snake River Plain Aquifer. 15th Ann. V.M. Goldschmidt Conference, Moscow, ID; May 20-25, 2005.
13. Delwiche, M. E., D. T. Newby, A. Wood, M. Bingham, R. L. Crawford, and F. S. Colwell. The effect of trichloroethylene on minimum energy requirement and gene expression in a nutrient limited methanotroph. Joint Intl. Symp. for

Subsurface Microbiology and Environmental Biogeochemistry, Jackson Hole, WY; Aug. 14-19-2005.

14. Strap, J. L., F. S. Colwell, and R. L. Crawford. Characterization of the eukaryotic diversity associated with planktonic and biofilm subsurface communities from the Snake River Plain Aquifer. Am. Soc. Microbiology Annual meeting, Atlanta, GA; June 5-9, 2005.

Graduate Students Trained

Two graduate students were trained during the project: J. L. Sebat, PhD and D. P. Erwin, MS.

Postdoctoral Scientists Trained

One postdoctoral scientist received experience during the project: Dr. Janice L. Strap

SPECIFIC ACCOMPLISHMENTS

1.0 Specific Accomplishment 1: find molecular markers of the methane monooxygenase (MMO) particulate form (pMMO) within the Test Area North (TAN) site of the Eastern Snake River Plain Aquifer (ESRPA) in Idaho

In this study we used primers specific for conserved regions of the *pmoA* gene of methane monooxygenase (MMO) particulate form (pMMO), the *mmoX* gene of MMO soluble form (sMMO), and the *amoA* gene of ammonia monooxygenase (AMO). We used the polymerase chain reaction (PCR) to amplify a variety of genes from whole community (metagenomic) DNA obtained directly from the aquifer. Based on these functional gene analyses, bacteria related to the type II methanotroph *Methylocystis* dominated both free-living and attached ESRPA communities. Type I sequences that grouped nearest to *Methylobacter* and *Methylomonas* were only obtained from the free-living community. Sequences similar to the *amoA* gene associated with ammonia-oxidizing bacteria (AOB) most closely matched a sequence from the uncultured bacterium BS870 but showed no substantial alignment to known, cultured AOB. Together these studies detected numerous sMMO-like gene sequences previously associated with high rates of TCE co-metabolism in both the free-living and the attached communities thus suggesting a metabolic potential of the indigenous populations to carry out natural attenuation.

These data are reported in: Erwin, D. P., I. K. Erickson, M. E. Delwiche, F. S. Colwell, J. L. Strap, and R. L. Crawford. 2005. Diversity of oxygenase genes from methane- and ammonia-oxidizing bacteria in the Eastern Snake River Plain aquifer. ***Appl. Environ. Microbiol.*** 71:2016-2025. Please refer to this document for detailed figures, tables, and literature citations. This work was supported in full by DOE EMSP award number DE-FG07-02ER63500.

1.1 Approach

The objective of this study was to use molecular techniques to examine the presence and diversity of oxygenase genes from methane-oxidizing bacteria (MOB) and ammonia-

oxidizing bacteria within a pristine region of the TAN site of the Idaho National Environmental and Engineering Laboratory (since renamed the Idaho National Laboratory), focusing on MMO and AMO genes of the total microbial community (free-living and attached microorganisms). This information will assist DOE in understanding the responses of the natural microflora of pristine aquifers to the appearance of contaminants such as trichloroethylene (TCE). Beyond simply confirming the presence of genes that enable TCE cometabolism, our aim was to assay the diversity of these genes as a first step in understanding how different versions of the same class of degradative enzymes might collectively contribute to TCE cometabolism.

1.2 Materials and Methods

1.2.1 Sampling. The microbial communities present in the ESRPA were sampled in the fall of 2003. Two wells, ANP-9 and ANP-10, located in a pristine region of the aquifer proximal to the TAN site's TCE-contaminated plume were used for this study. The close proximity of the wells to each other (ca. 200 m) suggests that their respective water chemistries and microbial community structures should be similar. Previous analysis indicates comparable water chemistry characteristics between the two wells and values that are typical of the ESRPA. For example, selected water chemistry values for ANP-9 and this region of the aquifer in general include temperature of 14.1-14.5°C, a pH of 7.2-7.9, a dissolved oxygen level of 7.5-8.4 mg L^{-1}, a dissolved organic carbon level of 1.5-2.6 mg L^{-1}, and a methane concentration of 72 nmol L^{-1}. Well ANP-9 was used in the collection of free-living cells, while well ANP-10 was used for attached community sampling.

The free-living community in this study was defined as all cells freely dispersed within the aquifer groundwater or only loosely associated with the basalt. The attached community consisted of all cells that colonized the surfaces of basalt substrates that were incubated in well ANP-10. Methods were designed to facilitate sampling of each community by using existing wells.

A KrosFlo Pilot hollow-fiber, tangential flow filtration system with a DynaFibre microporous membrane (0.2 μm pore size and 3.9 m^2 in surface area (Spectrum Labs, Rancho Dominguez, CA) was employed for the collection of free-living organisms from ANP-9. Prior to the initiation of filtration, the well was flushed with three well volumes (one volume equal to the standing water in the well) to remove built-up sediments and debris from the casing. Approximately 13,300 L of aquifer water was then passed through the hollow-fiber apparatus and concentrated to approximately 3 L. The concentrate was collected by reversing the pump motor and directing the retentate into sterile 1-L containers. This concentrate was then stored at -80°C and thawed at room temperature prior to being used.

The attached communities were sampled using substrate columns. These columns consisted of crushed basalt chips (4-7 mm sieve size) housed in a 76 cm long nylon mesh tubing (inside diameter, 2 cm; mesh size, 0.25 cm; InterNet, Inc., Minneapolis, MN). Prior to use, the columns were rinsed with deionized water to remove fine particles and then autoclaved (60 min) on three successive days. The sterile columns were then

suspended in the well to a depth of 80 m for 6 months. At the end of this period, the columns were removed and the attached microbial mass was harvested.

1.2.2 Metagenomic DNA isolation. An UltraClean Mega Prep soil DNA kit (MoBio Laboratories, Solana Beach, CA) was used to harvest the attached community metagenomic DNA from the basalt chips. Four 50-mL polypropylene tubes were filled with 40 g of the substrates and secured horizontally in a rotating incubator (Innova 4300; New Brunswick Scientific Co., Edison, NJ) at 200 rpm at 60°C for 30 min to remove bacterial biomass from the basalt substrates. Metagenomic DNA was recovered as per the manufacturer's protocol. After recovery, the extractions were pooled and concentrated by using a Microcon YM-30 ultrafiltration centrifugal device (Millipore, Billerica, MA).

Free-living cells were recovered by filtering 50 mL of the concentrated material through a 25-mm (0.20 μm-pore-size) Nuclepore polycarbonate membrane filter (Whatman, Clifton, NJ). DNA was then extracted from the filtered material by use of the UltraClean soil DNA isolation kit from MoBio. Briefly, the filter was placed into the bead solution and vortexed on high for 10 s to disperse the filtered material. The tubes were then placed in a 65°C water bath for 30 min. The purification of the lysed material was performed as per the manufacturer's protocol. This method was performed on three additional aliquots so that a total concentrate volume of 200 mL was processed. The resulting DNA from each extraction was pooled and stored at -20°C. DNA recovered from the substrate-attached bacteria by the modified MoBio Mega Prep method resulted in 140 μl of DNA at a concentration of 24 ng μL^{-1} ($A260/A280 = 1.5$).

1.2.3 Whole-genome amplification. Purified DNA from both communities was amplified by using the Repli-*g* whole-genome amplification (WGA) kit (Molecular Staging, New Haven, CT) to produce two additional community archives, W9 and W10. Briefly, the kit utilizes Φ29 polymerase in conjunction with exonuclease-resistant random hexamer primers to isothermally amplify DNA 104 to 106-fold with minimal bias. Based on multiple-displacement amplification technology, exponential amplification occurs through a hyperbranching mechanism. Long products, averaging 12 kb in length, are obtained due to the high processivity rate of 70,000 nucleotides incorporated for each primer binding event. Φ29 also exhibits the highest fidelity rate of any known polymerase, with an error rate of only 1 in 10^6-10^7 nucleotides incorporated.

For each sample, 10 ng of genomic DNA was chemically denatured by mixing with 2.5 μL of denaturation solution (40 mM KOH, 1 mM EDTA [pH 8.0]) and incubated at room temperature for 3 min. The reaction mixture was neutralized by the addition of 5 μL of neutralization solution (a 1:10 dilution of supplied solution B in distilled water). Amplification reactions were performed in 50-μL volumes containing 12.5 μL of 4 WGA PCR mix (as supplied by the manufacturer), 0.5 μL of DNA polymerase mix, 27 μL of sterile H_2O, and 10 μL of denatured sample. The reaction mixtures were then incubated at 30°C for 16 h. Amplification of the free-living DNA was only successful after two washes with distilled water followed by concentration with a Microcon filter device.

Amplified products were visualized by using 1% agarose gel electrophoresis stained with ethidium bromide. The amplified DNA was quantified by densitometry using Kodak 1D software (Kodak, Rochester, NY).

1.2.4 Functional gene PCR amplification. Three sets of PCR primers were used in this study to elucidate the oxygenase genes associated with methane- and ammonia-oxidizing bacteria from each community (free-living and attached) and WGA material (W9 and W10) (Table 1). To date, no single primer set has been demonstrated to amplify *mmoX* genes from all sMMO-producing bacteria.

Table 1. Primer sequences used for the molecular characterization of microbial communities derived from the ESRPA

Primer	Sequence (5' to 3')	Target
A189	GGNGACTGGGACTTCTGG	*amoA/pmoA*
A682	GAASGCNGAGAAGAASGC	*amoA/pmoA*
mb661	CCGGMGCAACGTCYTTACC	*pmoA*
mmoXA	ACCAAGGARCARTTCAAG	*mmoX*
mmoXB	TGGCACTCRTARCGCTC	*mmoX*

Primers that targeted the *mmoX* gene and the *pmoA* gene specifically were used, as was a set that allowed the co-recovery of *amoA* gene fragments along with *pmoA*. All PCR products were verified on 1% agarose gels stained with ethidium bromide and visualized using UV light. Images were captured and recorded by using Kodak 1D software. Reagents and enzymes were purchased from Fisher Scientific (Hampton, NH) unless otherwise specified. Oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA) with standard desalting.

pMMO diversity within the communities was assessed through the use of primers designed by others (Table 1; see Erwin et al. 2005). Both groups utilized the same forward primer, A189, in conjunction with one of two reverse primers. The first reverse primer, A682, was designed to target a highly conserved region in both the *pmoA* and *amoA* genes. These genes are believed to encode the active site containing the 27-kDa polypeptide subunit of both the MMO and the AMO. The A189-A682 primer set has been shown previously to amplify a 525-bp internal section of both *pmoA* and *amoA*. The second reverse primer, mb661, has been shown to preferentially amplify only an internal section of approximately 470 bp from the *pmoA* gene. PCR amplifications with both primer sets were performed in 50-µL volumes containing a 200-µM concentration of each deoxynucleoside triphosphate, 1x PCR buffer, 10 µg of bovine serum albumin, 1.2 mM $MgCl_2$, 10 pmol of each forward and reverse primer, 2.5 U of *Taq* polymerase, and 20 ng of DNA template. Reaction mixtures were subjected to initial denaturation for 5 min at 94°C followed by 30 cycles of 57°C for 1 min, 72°C for 1 min, and 94°C for 1 min. The final step consisted of 57°C annealing for 1 min, followed by 10 min of elongation at 72°C.

The primer set mmoXA and mmoXB was used to amplify an approximately 1,230-bp fragment from the *mmoX* gene. This primer set is designed to target the region of *mmoX*

that encodes the conserved fragment of the α-subunit of the hydroxylase component of sMMO. PCR amplifications were performed in 50μL volumes containing 200-μM concentrations of each deoxynucleoside triphosphate, 1x PCR buffer, 10 μg of bovine serum albumin, 1.2 mM $MgCl_2$, 10 pmol of each forward and reverse primer, 2.5 U *Taq* polymerase, and 20 ng of DNA template. Reaction mixtures were subjected to an initial denaturation of 5 min at 94°C followed by 30 cycles of 60°C for 1 min, 72°C for 1 min, and 94°C for 1 min. The final step consisted of 60°C annealing for 1 min, followed by 10 min of elongation at 72°C.

PCR of a metagenomic DNA template isolated from an undefined groundwater methanotroph enrichment that had been grown on methane was used as a positive control for each primer set.

Following amplification, all PCR products were purified by a Wizard spin column PCR cleanup (Promega, Madison, WI). Some products required additional purification by agarose gel electrophoresis in order to obtain clones of the proper insert size. This procedure was done by loading approximately 1 μg of PCR product into 1% SeaPlaque GTG LMP agarose (Cambrex Bio Sciences, Baltimore, MD) followed by extraction via Qiaex II gel extraction (Qiagen, Valencia, CA).

Archives were coded based on the well numbers used in the sampling (9 and 10 for free-living and attached communities, respectively) and the primer sets used in the PCR amplification. For instance, the archive produced from free-living cells isolated from well ANP-9 and amplified with the PCR set A189-A682 was coded as 9-682. Archives produced from whole-genome-amplified DNA were designated with a "W" preceding the well identifier.

1.2.5 Functional gene fragment library production. PCR amplicons for each sample were cloned by using the TOPO TA cloning kit for sequencing (version K) (Invitrogen, Carlsbad, CA). Briefly, 4 μL of each fresh PCR product was incubated with 1 μL of salt solution (1.2 mM NaCl, 0.06 M $MgCl_2$) and 1 μL of pCR4-TOPO vector. Electrocompetent Top10 *Escherichia coli* cells were electroporated at 2.4 kV in 0.2 cm cuvettes with 2 μL of the cloning reaction mixture. Positive transformants were selected by spread-plating onto Luria-Bertani agar plates with either kanamycin (50 μg mL^{-1}) or ampicillin (60 μg mL^{-1}) used as the selective agent and incubated overnight at 37°C. Kanamycin was used when direct PCR amplicons were cloned, whereas ampicillin was used when the amplicon was purified by gel electrophoresis.

Ninety-six colonies were isolated for each library archive and inoculated into 96-well plates containing 200 μL of Luria-Bertani broth and 50 μg mL^{-1} of kanamycin amended with 1x Haugness buffer (4% glycerol, 3.6 mM K_2HPO_4, 1.3 mM KH_2PO_4, 2 mM trisodium citrate, and 1 mM $MgSO_4$ at pH 7.5) as a cryoprotectant. These plates were then incubated in a GeneMachines HiGro microtiter-plate orbital shaker (Genomic Solutions, Ann Arbor, MI) for 18 h at 420 rpm and 37°C with an airflow rate of 4 standard L min^{-1} per chamber. Ten wells were randomly sampled for insert verification by PCR using vector-specific flanking primers (M13f and M13r supplied with the kit).

Amplicons were visualized on 1% agarose gels pre-stained with ethidium bromide. After verification, plates were sealed with aluminum tape and stored at -80°C for future use.

1.2.6 Restriction fragment length polymorphism (RFLP) analysis. Plasmid DNA was isolated from the 96-well cultures using the Montage Plasmid Miniprep96 kit (Millipore). Cultures were prepared by inoculating one round-bottom, 96-well cell culture block (supplied with the kit) containing 1 mL of Terrific broth with 50 µg of kanamycin ml^{-1} per well for each archive. Blocks were incubated in the HiGro incubator for 24 h under the conditions described above. Cells were harvested by centrifugation at 1,500 x g for 5 min with subsequent removal of supernatant by decanting. The blocks were sealed and pellets were frozen at -20°C prior to processing.

Plasmid DNA was isolated from the cell pellets as per the manufacturer's full lysate protocol. Five wells were selected from each plasmid preparation and DNA quantified by measuring the absorbance at 260 nm.

RFLP patterns were produced for each clone by digestion of 1 µg of plasmid DNA with 5 U each of MspI and HinPII (New England Biolabs, Beverly, MA) in a final 50-µL reaction mixture volume. Reaction mixtures were incubated at 37°C for 2 h, followed by enzyme inactivation at 65°C for 10 min. The manual grouping of fragmentation patterns was performed by analyzing 10 µl of each digest on 4% NuSieve 3:1 agarose (Cambrex Bio Science). Gels were stained with ethidium bromide and rinsed with deionized H_2O prior to imaging. One representative from each RFLP group was chosen for further analysis.

1.2.7 Sequencing and phylogenetic analysis. Sequencing of insert DNA was performed on an ABI Prism model 3730 sequencer (Applied Biosystems, Foster City, CA) by using BigDye version 3 chemistry. Contiguous sequences were produced by using ContigExpress software from the Vector NTI suite, version 9. Nucleotide-nucleotide BLAST (blastn) was used to search GenBank for nearest relative sequences. BLAST results and representatives for each archive were aligned by using AlignX software from the Vector NTI suite, version 9.

Edited alignments were evaluated with the maximum likelihood method using the PAUP package (version 4.0b10; Sinauer Associates, Sunderland, MA). Evolutionary distance calculations were generated using the model best describing the data for each tree as determined using the DT-ModSel program. Confidence estimates for the nodes within phylogenetic trees were performed by bootstrap analysis (100 replicates). Trees were visualized with TreeView software, version 1.6.6.

1.3 Results

1.3.1 Sampling. Free-living cell collection produced 3 L of concentrate at 4.2 x 10^6 cells mL^{-1} (1.26 x 10^{10} cells in total) as determined by direct microscopic counts of 4',6-diamidino-2-phenylindole (DAPI)-stained cells. After taking into account a 1:4,433 dilution factor, these results suggest a concentration of approximately 10^3 cells ml^{-1} in the original groundwater. This number agrees with previous cell density estimates from this

same well. Biomass from the substrate columns was not quantified. Metagenomic DNA isolated from sampling of free-living cells by the heat-lysis method used in conjunction with the MoBio kit produced a total of 160 µl of DNA at 12.5 ng μL^{-1} (*A*260/*A*280 = 1.61). The DNA was ~25 kb in size.

1.3.2 Whole-genome amplification. In an effort to assess the applicability of using WGA as a means to generate additional metagenomic template DNA for analyses, archives generated from amplified and unamplified templates were compared. WGA samples resulted in quantities of 550 ng μL^{-1} and 800 ng μL^{-1} (quantified by densitometry) for the free-living and attached communities, respectively. These concentrations represent nearly 104-fold amplifications. Both amplified samples migrated between 20 and 25 kb when analyzed by agarose gel electrophoresis.

1.3.3 PCR amplification. DNA extracted from the free-living and attached communities resulted in PCR amplicons of the expected size for a given primer pair. The only exception was the absence of a PCR product from DNA isolated from the attached community with the *mmoX*-specific primers. These primers were also unsuccessful at amplifying the targeted gene fragment from WGA attached community DNA. In some instances, the products did not resolve as single, isolated bands even after PCR optimization.

1.3.4 Gene fragment library production. Initially, we found that high transformation rates could be obtained, although a large percentage (30-50%) of those clones contained nonspecific inserts. In order to increase the percentage of clones containing the proper insert size, it was necessary to gel-purify the PCR product prior to cloning. This was the case even in the absence of other visible contaminating DNA, such as multiple bands or primer dimers. After gel purification, the percentage of clones not containing inserts of the proper size dropped well below 10%. In general, DNA from all communities transformed at a very low efficiency compared to positive controls supplied with the kit, though some efficiencies were far lower than others. The reason for these differences remains unclear, but we speculate that it may be due to the nature of the oligonucleotide primers used in the PCR. In most cases, multiple cloning and transformation reactions were necessary to produce 96 colonies for the archives.

1.3.5 RFLP and phylogenetic analysis. A total of 96 clones were screened for each of the 10 functional gene archives. Initially, many RFLP patterns were observed in all archives, and a majority of those groups contained only a single representative. Subsequent sequencing revealed that many of these patterns were a result of false-positive clones that did not contain the proper insert. These RFLP groups were discarded from the archives. Figure 1 from Erwin et al. (2005) shows a representative gel containing multiple RFLP patterns from 9-661, 9-682, W9-661, W9-682, and W10-682. Archives produced with primer set A189-A682 resulted in a combined total of 327 clones. From these clones, only type II methanotroph and AOB sequences that corresponded to α-*Proteobacteria* and β-*Proteobacteria*, respectively, were found (Figure 2 from Erwin et al. 2005). Archive compositions of WGA and unamplified communities were similar except that group I uncultured (*pmoA*) sequences were obtained with less frequency in

the WGA free-living community archive than in the unamplified archive, while *Nitrosospira* spp. (*amoA*) were represented to a much greater degree in the WGA-derived archive than in the unamplified archive. Interestingly, this difference was not observed for the attached community.

A comparison of the free-living and attached communities showed differences between the two communities. Group II uncultured (*pmoA*) sequences were more prevalent in the attached community than in the free-living community. Furthermore, representative sequences from group I and III uncultured (*pmoA*) bacteria and *Nitrosospira* were not found in the attached community archive (Table 2 from Erwin et al. 2005). Archive 9-682 produced a total of 83 clones from which 13 unique RFLP patterns were obtained (Figure 2 from Erwin et al. 2005). Just over half (59%) of the clones were found to contain oxygenase genes that group within α-*Proteobacteria*, while the remainder (41%) were from subclass β-*Proteobacteria*. BLAST results for groups 1, 3-7, 11, 13, 16, and 19 revealed sequences closely (95-98%) related to those of previously identified *pmoA*/*amoA*-containing bacteria. However, groups 2, 9, and 17 showed only 88% sequence identity to the oxygenase genes of uncultured bacterium M84-P36. These three groups branched separately from M84-P36 and formed a distinctive cluster with W9-682 7 (Figure 2 from Erwin et al. 2005). W9-682 produced a total of 91 clones that resulted in ten unique RFLP patterns of which 14% were α-*Proteobacteria* and 86% were β-*Proteobacteria* (Figure 2 from Erwin et al. 2005). All groups, with the exception of group 7, yielded high sequence identity (94-99%) to previously identified *pmoA*/*amoA*-containing bacteria similar to those in 9-682. Group 7, like groups 2, 9, and 17 of 9-682, had the highest identity value of 88% for the uncultured bacterium M84-P36. Group 11 formed a distinct branch within the tree.

Archives 10-682 and W10-682 were highly similar in composition and showed less oxygenase diversity than the free-living community archives, as only five RFLP patterns were seen for 10-682 and six patterns for W10-682 (Figure 2 from Erwin et al. 2005). Both archives contained a large majority (89% and 97%, respectively) of β-*Proteobacteria*, of which nearly all showed high levels of similarity to oxygenase genes of the uncultured bacterium gp22 and are distantly related to the AOB *Nitrosospira* (43 of 45 nucleotides were identical). Two of the groups of α-*Proteobacteria*, 10-682 4 and W10-682 10, clustered near the MOB *Methylocystis* sp. 42/22, while the third group of W10-682 15 clustered near other uncultured α-*Proteobacteria*.

Archives produced from primer set A189-mb661 resulted in a combined total of 317 clones. An overall high percentage of α-*Proteobacteria* was found in each of these archives. Archive 9-661 and W9-661 contained 97% and 77% *Methylocystis*, respectively, and both 10-661 and W10-661 were made up of 100% *Methylocystis*-like sequences (Table 3 from Erwin et al. 2005). Archives 9-661 and W9-661 were very similar in composition, with a high percentage of each corresponding to *Methylocystis* sp. (Figure 3 from Erwin et al. 2005). Seven RFLP patterns were seen in archive 9-661, whereas W9-661 resulted in 13 patterns. Four of the groups, 9-661 9, W9-661 5, W9-661 11, and W9-661 13, branched with the type I methanotroph *Methylomonas* sp., though they formed a divergent cluster. Another type I methanotroph *Methylobacter* sp. was distantly related

(81% sequence identity) to the groups 9-661 3 and W9-661 12. The nearest identity of these two groups is to the uncultured bacterium LP20 (89%), though the sequences were divergent enough to branch separately.

Archives 10-661 and W10-661 each produced six unique RFLP patterns and were similar in composition (Figure 3 from Erwin et al. 2005). All of the clones screened in each archive were highly related to oxygenase genes of *Methylocystis*, with 98% sequence identities in most cases and 95% sequence identities in all cases. These clones clustered very closely to the *Methylocystis* sp. used as the reference in the phylogenetic tree. No type I methanotrophs or AOB were found.

Archives produced from primer set mmoXA-mmoXB resulted in a total of 143 clones. Of these clones, 12 RFLP patterns were produced from 9-mmoX, and seven patterns were produced from W9-mmoX. Overall, both archives resulted in similar compositions, with *Methylocystis*-like oxygenase sequences resulting in 95% and 100% of the total clones for 9-mmoX and W9-mmoX, respectively (Table 4 from Erwin et al. 2005). Phylogenetic analysis of the two archives revealed three distinct clusters, two closely related to *Methylocystis* sp. and the other to *Methylomonas* sp. Type I methanotrophs found only in 9-mmoX (groups 5, 8, and 11) were related to *Methylomonas* sp. LW15 (91% sequence identity). These three groups collectively account for 5% of the overall archive. Type II methanotrophs represented in 9-mmoX include 87% *Methylocystis* sp. IMET 10486 (groups 1, 2, 4, and 14), 7% *Methylocystis* sp. 51 (groups 3, 13, 15, and 16), and 1% *Methylocystis* sp. SE12 (group 12). Type II methanotrophs represented in W9-mmoX include 68% *Methylocystis* sp. SE12 (groups 1, 5, and 7), 30% *Methylocystis* sp. IMET 10486 (groups 6, 10, and 12), and 2% *Methylocystis* sp. 51 (group 14) (Figure 4 from Erwin et al. 2005).

1.4 Discussion

1.4.1 Sampling. It has been reported that most bacteria in an aquifer are attached to the aquifer solids, and only a small fraction exist as free-living cells. The sampling of attached biomass from aquifer boreholes is usually accomplished through the collection of core samples. This method can be expensive, as well as challenging, depending on the aquifer to be sampled. In this study, we chose to use porous columns filled with material simulating the existing fractured basalt of the aquifer and incubated in a borehole in the aquifer.

The use of artificial substrata incubated in wells to sample attached communities close to the study site was described previously (see reference in Erwin et al. 2005); it has been found that the use of dialysis chambers containing basalt (similar to the substrate columns used here) resulted in the recovery of a greater range of microbial diversity compared to that from authentic core samples. Therefore, boreholes may not be ideal for collecting representative samples of microbes colonizing aquifer solids.

Free-living communities sampled by filtration of groundwater are also subject to biases. Due to the large volume of groundwater used in this study (13,300 L), it is likely that many different microbial niches were sampled, not all necessarily corresponding to the

free-living community. Sediment particles were present in the filtrate, and bacteria colonizing these sediment layers would have been extracted along with free-living cells. Other have observed that bacteria display a dynamic equilibrium between attached and free-living phases. These bacteria form biofilms as part of their normal life cycle that culminates with the release of free-living cells capable of colonizing new habitats. Thus, it is difficult to draw a distinct line between the attached and free-living communities. Though the communities may differ in overall composition, it is not surprising that we found no distinct disparity between the attached and free-living communities sampled. It is likely that free-living cells present in the free-living samples are also represented to some extent in the attached samples.

1.4.2 Whole-genome amplification (WGA). WGA technology is based on the unbiased amplification of limited DNA samples. This technology was able to amplify our starting DNA to produce a product equal in size to the starting template when visualized by agarose gel electrophoresis. PCR of WGA products also resulted in amplicons equal in size to unamplified samples. The differences noted between archives were most likely due not to the WGA process but rather to bias introduced by the PCR. Since multiple PCRs were not pooled, bias generated through a single PCR amplification may have skewed the resulting archives.

1.4.3 Functional gene library production and phylogenetic analysis. The primer sets A189-A682, A189-mb661, and mmoXA-mmoXB were used successfully as functional gene probes in the assessment of MOB and AOB in most of the communities tested. It is unclear why no amplification was seen from the attached and WGA attached DNA with the mmoXA-mmoXB primer set. Sequences were amplified from both attached and WGA-attached communities by using the primer sets A189-A682 and A189/mb661, and these were closely related to sequences of the sMMO-producing MOB *Methylocystis*.

An explanation for our observations may lie in the distinct difference in the species of *Methylocystis* found in free-living versus attached communities. None of the sequences amplified from the free-living samples with A189-A682 and A189-mb661 was found in the mmoXA-mmoXB archives and vice versa. It appears that the three primer sets used exhibit specificity towards different *Methylocystis* spp. This fact supports the conclusion that the methanotroph species targeted by mmoXA-mmoXB were not present in the attached samples in sufficient numbers to be detected.

The number of available *mmoX* sequences on which to base primer design is small. The primer set *mmoX* f882-*mmoX* r1403, widely used for *mmoX* amplification, is based on sMMO gene clusters from only two methanotrophs. The mmoXA-mmoXB primer set, which was designed five years after the introduction of *mmoX* f882-*mmoX* r1403, is based on only six known full *mmoX* sequences. It may be that genes found to be conserved in cultivable organisms that were used to design the primers may not be sufficiently similar to genes in the environment, and this possibility should be considered when interpreting data generated by such primer sets. To date, no single primer set has been demonstrated to amplify *mmoX* genes from all sMMO-producing bacteria.

As expected, the primer set A189-A682 amplified sequences corresponding to both *pmoA* and *amoA* from all four communities. The abundance of *amoA*-related sequences compared to *pmoA* found in all four archives except 9-682 suggests that nitrifying bacteria may dominate the samples. Of particular interest is the presence of the large subset of clones that cluster with the uncultured bacterium M84-P36. These clones make up a total of 35% of the archive and are only distantly related to other uncharacterized *pmoA* clones, suggesting the presence of a potentially significant novel group of *pmoA*-carrying methanotrophs. The proportion of sequences in clone libraries, however, cannot be used to accurately predict the relative abundance of the AOB or the novel *pmoA* group in the natural communities due to biases that are inherent in PCR amplification and the number of sequences analyzed. Such proportions, however, are useful for comparisons of archive data within a given data set.

Characterization of the communities with *pmoA*-specific primers (A189-mb661) suggested that bacteria related to the type II methanotroph *Methylocystis* dominate free-living as well as attached methanotroph communities. The free-living communities (amplified and unamplified) also showed the presence of type I methanotrophs related to *Methylomonas* sp. LW15, a known sMMO-containing bacterium. The small percentage of clones from archives 9-661 and W9-661 found to be related to the genus *Methylobacter*, which has not been shown to produce sMMO, suggests that a broad range of niches may have been sampled through the collection of free-living cells.

It has been proposed that under low oxygen, high methane conditions, growth of type II methanotrophs is favored over growth of type I methanotrophs. The analysis of groundwater sampled from well ANP-9 showed a significant concentration of methane but also revealed an oxic environment. It is expected that an environment such as this should favor the growth of type I methanotrophs. The presence of copper, however, is another key factor responsible for regulating the growth of type I and type II methanotrophs. Copper is necessary for pMMO activity and plays a key role in the regulation of pMMO and sMMO expression. Under low-copper conditions, bacteria with the ability to produce sMMO are able to out-compete those able to produce only pMMO. This mechanism is due to the fact that the active pMMO enzyme contains approximately 12-15 copper atoms mol^{-1}. While the exact nature of these Cu^{2+} ions in the function of pMMO is not clear, they are postulated to play a role in the active site of the protein. Samples of groundwater from ANP-9 show that copper levels are below the detection limit of 1.0 $\mu g\ mL^{-1}$. Taking into account the conditions present in the wells sampled (the presence of dissolved methane and low copper concentrations), the community composition observed in this study is not surprising.

Our results, obtained through the molecular assessment of oxygenase genes, demonstrate that in both the attached and free-living communities, the majority of the methanotrophic populations present show a high degree of similarity to other type II methanotrophs. The natural attenuation of TCE observed at the TAN site may be due in part to the sMMO-producing communities described here. However, we did not attempt to enumerate methanotrophs from either free-living or attached communities. Because of this fact, the

organisms represented in this study constitute an undetermined percentage of the overall bacterial population.

The presence of sMMO-producing bacteria alone does not account for the rate of TCE natural attenuation observed at the TAN site. It was our goal to provide evidence that the *in situ* bacterial populations within the pristine aquifer have the potential for TCE cometabolism. Future work to quantify type II methanotrophs specifically and monitor the production of sMMO is necessary in order to correlate the presence of these bacteria with the TCE attenuation rate observed at the TAN site within the contaminated plume.

****

2.0 Specific Accomplishment 2: determine eukaryotic diversity in planktonic and biofilm populations of the Snake River Plain Aquifer

Eukaryotic microorganisms are an important component of microbial communities in aquifers, yet phylogenetic information with respect to their diversity within this habitat is scarce. Using molecular methods, we characterized the eukaryotic microbial diversity of planktonic and biofilm communities within the oxic, fractured basalt Snake River Plain Aquifer (SRPA). Planktonic microorganisms were collected by filtration of pumped groundwater, and biofilms were obtained by colonization of basalt chips housed within substrate columns suspended in the aquifer. 18S rDNA sequences were placed into operational taxonomic unit (OTU) groups with ≥99% 18S rDNA similarity. The planktonic community was composed of fungi (49%), stramenopiles (32%), and choanoflagellida (19%), while the biofilm community was comprised of choanoflagellida (70%), cercozoa (16%), and fungi (14%). Three novel fungal clades, five novel stramenopile clades, and seven novel choanoflagellate sequences were identified. The planktonic community harbored greater diversity than the corresponding biofilm as indicated by the reciprocal of Simpson's index (1/D) and the Shannon-Weiner index. Non-parametric estimation of the species richness by Chao1 revealed that fewer species were expected in the biofilm population (five species) than in the planktonic community (30 species). The ∫-LIBSHUFF program provided evidence that the planktonic-derived OTUs were significantly different ($p = 0.0001$) from those of the biofilm. This was confirmed by F_{ST} analysis ($F_{ST} = 0.08497$, $p<0.0001$). While planktonic protists have been previously described in aquifers, to our knowledge this is the first report of planktonic and biofilm-associated eukarya characterized from an aquifer by molecular methods.

These data have been submitted for publication (Strap, J. L., F. S. Colwell and R. L. Crawford. In Review. Eukaryotic diversity in planktonic and biofilm populations of the Snake River Plain Aquifer. *Microbial Ecology*). A copy of this submission is attached and contains relevant tables, figures, and references. This work was supported in full by DOE EMSP award number DE-FG07-02ER63500.

2.1 Approach

Molecular methods have been extensively used to investigate prokaryotic diversity in a wide range of environments; however, molecular investigations of eukaryotic populations are more recent and have not yet been as extensive. The reason for this is in part due to a well-established ultra-structure and morphology-based eukaryote taxonomy. However, morphological classification can be complicated by the sheer number, diversity and size range of many species within an environmental sample. The molecular studies of eukaryotes that have been conducted have focused on soils, desert rock varnish, rivers, solar salterns, marine environments, anoxic sediments, and acid mine drainage. There have been relatively few prior reports of eukaryotic diversity within aquifers, with the work that does exist consisting of visual/microscopic and/or culture-based evaluations. Microeukaryotes have a direct impact on the ecology of the aquifer due to their important role in the food web of this environment. For example, the consumption of bacteria and archaea by protists affect the abundance of these organisms and thereby play a principal role in nutrient cycling.

Studies of biofilms have historically focused on prokaryotes and therefore, knowledge of substrate-associated eukaryotes is sparse. It is likely that microeukaryotes such as protists influence the biofilm architecture by their own presence in addition to their feeding on substrate-associated bacteria and archaea. Population dynamics of protozoa in biofilms has been investigated *in vitro*, but to our knowledge a molecular characterization of biofilm-associated microeukaryotes sampled from deep, oxic aquifers has not been conducted.

The rather limited number of currently available 18S rRNA sequences negatively affects the ability to design appropriate primers for the culture-independent assessment of microbial diversity. One way this limitation can be overcome is to add to the inventory of microeukaryotic phylotypes.

Investigations into the diversity of microeukaryotes within deep aquifers will give insight into key relationships between species richness and organic contamination. This is particularly important in areas such as the Test Area North (TAN) site of the Idaho National Laboratory (INL) in southeast Idaho where a trichloroethylene (TCE) plume contaminates the Snake River Plain Aquifer (SRPA) that underlies the site. The concentration of dissolved TCE at the site is decreasing by natural attenuation, and this has positive, economic impact for remediation of the site. The participation of eukarya in intrinsic bioremediation should not be overlooked. To fully understand the process of natural attenuation, knowledge of the indigenous eukarya in pristine regions is required. Here we report results of our studies of the planktonic and biofilm microeukaryotic diversity found within a pristine region of the oxic, fractured basalt SRPA. We found that the planktonic community was more diverse than the biofilm community. To our knowledge this is the first report of an assessment of eukaryotic diversity of an aquifer environment by molecular methods.

2.2 Materials and Methods

2.2.1 Sampling. Two wells in a pristine region of the Snake River Plain Aquifer near the Test Area North (TAN) site at the Idaho National Laboratory were used for sampling. The planktonic community was sampled from well ANP-9 by concentrating the cells present in 13,300 L of aquifer water to approximately 3 L by employing a KrosFlo Pilot hollow-fiber, tangential flow filtration system with a DynaFibre microporous membrane (0.2 μm pore size and 3.9 m^2 surface area; Spectrum Labs, Rancho Dominguiz, CA). The well was flushed with three well volumes (one volume equal to the standing water in the well) to remove built-up sediments and debris from the casing prior to filtration. The concentrate was collected by reversing the pump motor and directing the retentate into sterile 1 L containers which were subsequently stored at -80°C and thawed at room temperature prior to use. The biofilm community was sampled from well ANP-10 by incubating sterile basalt substrates (4-7 mm sieve size) housed in 76-cm-long nylon mesh tubes (ID, 2 cm; mesh size 0.25 cm; InterNet, Minneapolis, MN) in the aquifer for six months at a depth of 80 m. The wells chosen for the study are in close proximity (ca. 200 m) to each other. Previous analysis indicated comparable water chemistry between the two wells and that the values are typical of the SRPA. For example, a temperature of 14.1-14.5°C, pH of 7.2-7.9, dissolved oxygen level of 7.5-8.4 mg L^{-1}, dissolved organic carbon level of 1.5-2.6 mg L^{-1}, and methane concentration of 72 nmol L^{-1} have been reported.

2.2.2 Total DNA isolation. Removal of biofilm biomass from the basalt substrates was achieved by filling four 50-mL polypropylene tubes with 40 g of the basalt substrates and incubating at 200 rpm at 60°C for 30 min in a horizontal position in a rotating incubator (Innova 4300; New Brunswick Scientific Co., Edison, NJ).

Planktonic cells were recovered by filtration of the concentrated material through a 25-mm (0.2-μm pore size) Nuclepore polycarbonate membrane filter (Whatman, Clifton, NJ).

DNA from the planktonic and biofilm communities was isolated using the UltraClean™ Soil DNA Kit Mega Prep (MoBio Laboratories, Solana Beach, CA) following the manufacturer's instructions. This method has been shown to be effective for extraction of DNA from both prokaryotes and eukaryotes. Two independent extractions from each community were pooled for use as template. The resulting DNA was stored at -20°C.

2.2.3 Clone library construction. Eukaryotic 18S rRNA genes were amplified from total DNA template using NS3 (nu-SSU-0505-5'; 5'-GCAAGTCTGGTGCCAGCAGCC-3') and NS8 (nu-SSU-1769-3'; 5'-TCCGCAGGTTCACCTACGGA-3') primers. These primers were previously described for amplification of fungal rDNA with analysis of the primer sequences by BLAST; however, ARB [ARB (arbor) is a graphically oriented software program for nucleic acid sequence database handling and analysis (http://www.arb-home.de] revealed that this primer pair had a broad specificity including Metazoa, Viridiplantae, Fungi, Ichthyosporea, Acanthamoebidae, Haptophyceae, Cryptophyta, Rhodophyta, Stramenopiles, Cercozoa, Alveolata, Polycystinea, Diplomonadida, Apusomonadidae, Glaucocystophyceae, Choanoflagellida, Mesomycetozoa, and Epsilon-proteobacteria. Each PCR mixture (50 μL) contained 1x PCR buffer (Sigma-Aldrich, St. Louis, MO), 2.5 mM $MgCl_2$, 200 μM (each) deoxyribonucleoside triphosphate (dTTP,

dATP, dGTP, and dCTP), 1.0 µM (each) primer, and 20-50 ng template DNA and JumpStart™ *Taq* DNA polymerase (0.05 U; Sigma-Aldrich). PCR amplification was performed on an MBS thermocycler (ThermoHybaid, Franklin, MA) under the following conditions: initial denaturation of the template DNA at 95°C for 5 min, 25 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 60 s, and a final elongation step at 72°C for 10 min. The PCR products were visualized by electrophoresis on 1% (w/v) agarose gels. Positive controls consisted of reactions containing DNA from *Candida albicans* ATCC 90027 and *Aspergillus nidulans.* Negative controls consisted of PCR reactions containing no template DNA addition, *Archaeoglobus fulgidus* ATCC 49558, *Halobacterium salinarum* ATCC 19700 or *Escherichia coli* DNA as template. To minimize PCR bias, four independent PCR reactions were pooled. The combined reaction products were purified with a MoBio Ultraclean™ PCR Clean-Up™ kit (MoBio Laboratories Inc., Carlsbad, CA) according to the manufacturer's instructions. The amplicons were cloned using the TOPO TA Cloning® Kit for Sequencing (version k) (Invitrogen Corp., Carlsbad, CA). Clones were grown in 96 well format plates and plasmid DNA was purified using the Montage Plasmid Miniprep$_{96}$ Kit (Millipore, Billerica, MA)..

2.2.4 Restriction fragment length polymorphism (RFLP) analysis. RFLP patterns were produced for each clone by digestion of 1 µg plasmid DNA with 5 U of *Hae*III (Invitrogen) in a final reaction volume of 50 µL. Reactions were incubated at 37°C for 2 hr followed by enzyme inactivation at 65°C for 10 min. Fragmentation patterns were analyzed by agarose gel electrophoresis on 3.5% NuSieve 3:1 agarose (Cambrex Bio Science, Baltimore, MD). Gels were stained with ethidium bromide and visualized under UV transillumination. Kodak 1D Image analysis software (Kodak Eastman, New Haven, CT) was used to facilitate grouping of clones.

2.2.5 Sequence and phylogenetic analysis. One representative from each RFLP group was sequenced in both directions using T3 and T7 primers. Nucleotide sequence determinations of each representative clone were performed on an ABI Prism Model 3730 (Applied Biosystems, Foster City, CA) using Big Dye version 3 chemistry at the Laboratory for Biotechnology and Bioanalysis (School of Molecular Biosciences, Washington State University, Pullman, WA). Contiguous sequences were obtained using ContigExpress (Vector NTI suite version 9). Prior to comparative sequence analysis, vector sequences flanking the 18S rDNA inserts were manually removed. Sequences were analyzed for chimeras using the Ribosomal Database Project II CHECK_CHIMERA program and Bellerophon. Three putative chimeric planktonic sequences and one putative chimeric biofilm sequence were detected and removed from the analysis. Nucleotide-nucleotide BLAST (Basic Local Alignment Search Tool) was used to search GenBank for nearest relative sequences. BLAST results, representative eukarya sequences and RFLP representative clones were aligned using ClustalX and MUSCLE. Highly variable regions of the alignment were removed with Gblocks using parameters optimized for rDNA alignments (minimum length of a block was five; allowing gaps in half positions), leaving 1,307 informative positions. Edited alignments were evaluated with the Maximum Likelihood (ML) method using PAUP* (version 4.0b10; Sinauer Associates, Sutherland, MA) under the TrNef+G model of evolution, which was determined to be the model best describing the data as determined by the DT-

ModSel program. PAUP* was run on a cluster computer that has 130 dual Xeon processors and on a Silicon Graphics Octane 2 workstation. Confidence estimates for the nodes within the phylogenetic trees were performed by bootstrap analysis (1,000 replicates). Trees were visualized with TreeView software version 1.6.6. Maximum Parsimony (MP) bootstrap values were computed with PAUP by using a heuristic search method with a tree bisection-reconnection branch-swapping option with random taxon addition.

2.2.6 Rarefaction, clone coverage, and species richness estimation. Diversity coverage by the clone libraries was analyzed with DOTUR. Clone coverage was estimated as described by Mullins et al. (1995). The percent coverage (C) of the clone libraries was calculated according to the equation $C = [1-n_1/N] \text{ X } 100$ where n_1 is the number of unique clones and N is the total number of clones analyzed. The eukaryotic species richness was calculated for both the planktonic and biofilm communities using the nonparametric estimators ACE (abundance-based coverage estimator) and Chao1 using DOTUR.

2.2.7 Statistical comparison of coverage. The planktonic and biofilm 18S rDNA libraries were compared using the ∫-LIBSHUFF (http://www.plantpath.wisc.edu/fac/joh/S-libshuff.html) computer program to generate homologous and heterologous coverage curves from the 18S rDNA clone libraries. Sequences were randomly shuffled 999 times between samples prior to calculating the distance between the curves using the integral form of the Cramér-von Mises test statistic. The DNADIST program of PHYLIP (http://evolution.genetics.washington.edu/phylip.html) using the Jukes-Cantor model for nucleotide substitution was used to generate the distance matrix analyzed by ∫-LIBSHUFF. The ∫-LIBSHUFF program provides a statistical test for the null hypothesis that two 18S rRNA gene libraries are samples of the same community.

2.2.8 Comparison of phylogenetic diversity. The phylogenetic diversity between the planktonic and biofilm community was compared using *F* statistics (F_{ST}) calculated using the ARLEQUIN program. The F_{ST} test was used to compare the genetic diversity within each community to the total genetic diversity of the communities combined, using the equation $F_{ST} = (\theta_T - \theta_W)/\theta_T$, where θ_T is the genetic diversity for all samples and θ_w is the genetic diversity in each community. Statistical significance of F_{ST} was evaluated by randomly assigning sequences to populations and calculating F_{ST} for 1,000 permutations. An $F_{ST} \approx 0$ indicates that genetic diversity overlaps between the communities being compared. As the difference between the populations increases, the F_{ST} value also increases.

2.2.9 Diversity indices. Diversity indices utilized to describe and compare the biofilm and planktonic libraries for eukaryotic diversity were determined as described by Martin (2002). The reciprocal of Simpson's index (1/D) and the Shannon-Weiner index were used as general diversity measures and were calculated in DOTUR.

2.2.10 Nucleotide Sequence Accession Numbers. The sequences of the cloned inserts were deposited in GenBank under accession numbers DQ104580 – DQ104608.

2.3 Results

2.3.1 Planktonic community structure. Total DNA extracted from groundwater at well ANP-9 at the TAN site was used to construct a clone library of partial eukaryotic 18S rDNA genes. Initially, 89 planktonic clones were screened by RFLP analysis and grouped according to RFLP type. Three putative chimeras were removed from further analysis. Twenty-four unique phylotypes were obtained from the remaining 86 non-chimeric clones analyzed. The phylogenetic tree in Figure 1 (from Strap et al. in review and attached to this document) shows the relationship among the 24 planktonic phylotypes. The eukaryotic planktonic community was comprised mainly of fungi (49%) with the choanoflagellates and stramenopiles accounting for 19% and 32% of the clones analyzed, respectively (Figure 2 from Strap et al., in review). Two novel planktonic clades belonging to the fungi (clade FII, EPK BH1 and EPK B G4; clade FIII, EUKDPK65, EPK B B11, EPK B G7, and EUKDPK41) and one belonging to the choanoflagellida (EUKDPK38, EPK B E10, EPK B C12, EPK B C11, and EPK B C4) were observed (Figure 1 from Strap et al., in review). Four novel clades within the stramenopiles were also identified within the planktonic community: clade SI, represented by EPK B G5 and EPK B F7; clade SII, represented by EUKDPK45; clade SIII, represented by EUKDPK52, EUKDPK44, and EUKDPK69; clade SIV, represented by EPK B A10 and EPK B C3; and clade SV, represented by EPK B C10. Clade SIII was most closely affiliated with an uncultured clone obtained from Spain's 'River of Fire' (the Rio Tinto, an acidic stream flowing through the world's largest pyretic belt in southwestern Spain), while the other novel stramenopile clades did not affiliate with any known sequences.

2.3.2 Biofilm community structure. Total DNA recovered from colonized basalt chips after submersion for 6 months in well ANP-10 was used to construct a clone library of partial eukaryotic 18S rDNA genes. Eighty-one non-chimeric clones were screened by RFLP analysis. Of these clones, it was found that 44 sequences (representing 35 distinct RFLP groups) were determined to be archaeal in origin and were subsequently excluded from this study. From the 37 remaining eukaryal biofilm clones, five distinct RFLP patterns were observed. It is interesting to note that the NS3/NS8 primer set did not yield an amplicon when the archaea *Archaeoglobus fulgidus* or *Halobacterium salinarum* DNA was used as template. Furthermore, no archaeal sequences were amplified with these primers when the total DNA from the planktonic community was used as template. The phylogenetic tree in Figure 1 (Strap et al., in review) shows the relationship among the five biofilm phylotypes. A novel clade belonging to the choanoflagellida (EBF C D10 and EBF C E3) was observed. The eukaryotic biofilm was composed mainly of choanoflagellates (70%), fungi (14%), and cercozoa (16%, Figure 2 from Strap et al., in review).

2.3.3 Rarefaction analysis. Rarefaction analysis of the RFLP results obtained for both the planktonic- and biofilm-derived communities was performed to estimate the extent to which the eukaryal diversity of the two communities could be described by screening 86 and 37 non-chimeric clones, respectively. Rarefaction analysis results for the two clone libraries revealed that the number of planktonic clones analyzed was insufficient to describe the total extent of eukaryotic diversity since no obvious plateau was observed.

Thus, further analysis of 18S rDNA clones for the planktonic community would likely have revealed additional diversity. In contrast, the rarefaction curve for the biofilm clones did reach saturation, indicating that the diversity within the biofilm community had been adequately sampled. Percent coverage was determined to be 72% and 87% for the planktonic and biofilm libraries, respectively. The higher coverage of the biofilm is consistent with the lower diversity observed in this library.

2.3.4 ∫-LIBSHUFF comparisons of the planktonic and biofilm communities. To determine whether the planktonic and biofilm 18S rDNA libraries were samples of the same communities, ∫-LIBSHUFF analysis was performed. The ∫-LIBSHUFF comparisons of the planktonic and biofilm communities suggested that the two communities were derived from populations of different composition. ∫-LIBSHUFF analysis of homologous and heterologous coverage curves indicated that the planktonic and biofilm communities were significantly different from each other ($p = 0.001$). The heterologous coverage of the planktonic library by the biofilm library was significantly different ($p = 0.001$) from the homologous coverage of the planktonic library by itself thus indicating that some sequences from the planktonic library had no close relatives in the biofilm library. This is supported by visual inspection of the phylogenetic tree (Figure 1 from Strap et al., in review). Homologous coverage, a measure of similarity within a given library, was 100% for the biofilm library and 96% for the planktonic library. Heterologous coverage, a measure of the representation of a given library in another, was 43% for the biofilm compared to the planktonic and 17% for the planktonic compared to the biofilm at an evolutionary distance of 0.01 (99% 18S rDNA similarity OTU definition).

2.3.5 Species richness. Non-parametric estimation of species richness calculated by two different methods (ACE and Chao1) in DOTUR revealed that the planktonic community was more species-rich than the biofilm community (Table 1 from Strap et al., in review).

2.3.6 Phylogenetic diversity and comparison of diversity indices. Genetic differentiation among the sampled microbial communities was assessed using the F_{ST} test. If there is no effect of sample origin on phylogeny, the F_{ST} value should be close to zero. The more pronounced the effect of sample origin on phylogeny, the more the value of F_{ST} will deviate from zero. An F_{ST} value of 0.08497 ($p<0.001$) was obtained for the two communities demonstrating that the levels of genetic diversity were significant, implying that the within-community variation was different compared to the total diversity sampled when the two communities were combined. This is in agreement with the ∫-LIBSHUFF results. Since the choanoflagellates accounted for a significant portion of the biofilm population and were also found in the planktonic community and since it appeared that these sequences comprised distinct clades, it was of interest to determine how genetically related the planktonic and biofilm choanoflagellate sequences were to each other. An F_{ST} value of 0.50769 ($p<0.001$) was obtained for the comparison of the choanoflagellate sequences between the two communities indicating that the planktonic and biofilm populations were genetically discrete.

Various diversity indices were calculated for the planktonic and biofilm eukaryotic libraries (Table 1 of Strap et al., enclosed). The reciprocal of Simpson's index is sensitive

to the level of dominance in a community and indicated that the planktonic eukaryotic community (1/D index of 25) was five times more diverse than the biofilm community (1/D index of 5). The values for the Shannon-Weiner index, gene diversity, and total genetic variation show similar patterns (Table 1 of Strap et al., enclosed) with the planktonic community being more diverse than the biofilm community which is also reflected in the phylogenetic distribution. The planktonic library indeed contained representatives of more phyla than did the biofilm library (Figure 1 of Strap et al., enclosed).

2.4 Discussion

We have compared the eukaryotic diversity of both the planktonic and biofilm communities of the SRPA. Compared to prokaryotic diversity, comparatively little is known about the eukaryotes that are members of deep subsurface, oxic aquifer habitats. While planktonic and sediment-associated protists have been previously described in aquifers, to our knowledge, this is the first report of a molecular-based community assessment that includes biofilm-associated eukarya sampled from an aquifer and the first to apply species richness and phylogenetic diversity estimates of eukarya to an aquifer environment.

Rarefaction analysis was performed to determine the number of unique bacterial clones as a proportion of the estimated total diversity. The rarefaction curve for the planktonic-derived eukaryotic population did not reach a clear saturation indicating that further sampling of this clone library may have revealed additional diversity. In contrast, the rarefaction curve for the biofilm eukaryotic community reached a clear saturation suggesting this library had been adequately sampled and that further sampling would not likely reveal additional diversity. This is in agreement with non-parametric estimates for both communities (Table 1 from Strap et al., in review).

Comparison of all diversity measures (Table 1 from Strap et al., in review) combined with the F_{ST} analysis (0.08497, $p<0.0001$) showed that eukaryotic diversity in the planktonic fraction of the SRPA is much higher than the eukaryotic biofilm diversity. The two eukaryotic communities were also significantly different as determined by ∫-LIBSHUFF. These observations are consistent with the findings from previous studies on attached (biofilm) and planktonic prokaryotic communities in fractured rock aquifers that found that planktonic and attached communities in an acidic crystalline rock aquifer were distinct with respect to the biomass and the physiological capacities of the two fractions. In the SRPA, prokaryotic communities present in water pumped from well TAN-33 (located approximately 3 km from wells ANP-9 and ANP-10) were different from prokaryotic communities obtained from basalt substrates that were incubated in the well for eight months. Differences between planktonic communities in the attached (biofilm) and planktonic phases of an aquifer may be related to the differences that we observed in the eukarya present in these distinct phases.

The difference in diversity between the attached and planktonic communities may also have been due to differences in how the two communities were sampled. For planktonic samples, we concentrated the cells present in 13,300 L of groundwater. Given that the

basalts in the aquifer have a nominal porosity of 12%, this 13,300 L of groundwater would have been obtained from a subsurface volume of approximately 40,000 L (or 40 m^3). By comparison, the basalt substrates that were incubated for six months in ANP-10 in order to collect the attached biomass assayed a smaller volume of the aquifer.

Novel clades belonging to the fungi were observed for both the biofilm and planktonic communities: 1) clade FI (EUKDBF10) representing 27% (10/37) of the recovered biofilm sequences; and 2) clade FII (EPK B H1 and EPK B G4) and clade FIII (clones EPK B G7, EUK DPK 65, EUKDPK41, and EPK B B11) representing 7% (6/86) and 19% (16/86) of the recovered planktonic sequences, respectively. Novel choanoflagellates (EBF C E3 and EBF C D10; EUKDPK38, EPK B E10, EPK B C12, EPK B C11, and EPK B C4) were also observed in both communities. Novel clades belonging to the stramenopiles were observed only in the planktonic population: clade SI (EPK B G5 and EPK B F7) representing 8% (7/86), clade SII (EUKDPK45) representing 3% (3/86), clade III (EUKDPK52, EUKDPK44 and EUKDPK69) representing 10% (9/86), clade SIV (EPK B A10 and EPK B C3) representing 2% (2/86), and clade SV (EPK B C10) representing 7% (6/86) of the recovered planktonic sequences. It is noteable that clade SIII affiliated with an uncultured eukaryote clone isolated from Spain's River of Fire (uncultured eukaryote clone Rt5ii25; AY082983). This clone was suggested to be a member of a novel stramenopile lineage, and our results confirm this. Overall, 30% of the planktonic community was composed of novel stramenopile sequences. Massana et al. (2004) observed that novel stramenopiles were found in all major marine habitats and are important members of eukaryotic picoplankton. Our results suggest this is also true for oligotrophic aquifer environments such as the SRPA. We did not observe any stramenopile representatives within the biofilm population. These novel environmental sequences significantly expand the known extent of eukaryotic rDNA diversity within the deep subsurface. It is possible that some of these sequences may represent eukaryotes identified in previous culture or microscopic studies but are not represented by rRNA sequences in the database. Due to the paucity of available 18S rRNA sequence information, concrete conclusions with respect to the significance of these novel clades must be made with caution. It is interesting that when sequenced, 44 out of 81 eukaryotic biofilm-derived clones analyzed were determined to be archaeal in origin. Although archaea are present in the aquifer, this phenomenon was not observed with the planktonic-derived clones nor was it observed when two separate archaeal genomic templates were used as controls. The NS3/NS8 primer pair, allowing for a single mismatch, will amplify archaeal representatives.

The planktonic and biofilm eukaryal communities were found to be significantly different (Figure 2 and Tables 1 and 2 from Strap et al., in review). The fungi dominated the planktonic eukaryal community (49%) while the choanoflagellates were the most abundant group within the biofilm community (70%). Cercozoa (16%, Figure 2 and Table 2 from Strap et al., in review) were only found in the biofilm population and were represented by a single clone (EUKDBF47), which was most closely affiliated with an uncultured cercozoan clone (AY620270).

Choanoflagellates are generally considered to be significant members of marine and freshwater heterotrophic flagellate assemblages and are among the closest living single-celled relatives of metazoans and may represent a model taxon for investigating metazoan origins. That we found novel representatives within the choanoflagellates may contribute significantly to the investigation of metazoan origins. It is interesting that the planktonic choanoflagellate population are genetically distinct lineages from the biofilm choanoflagellates (F_{ST} = 0.50769, $p<0.001$). Whether or not this difference is due to selective feeding remains to be investigated. Heterotrophic flagellates such as the choanoflagellates have been recognized as consumers of planktonic and attached (biofilm) bacteria. Choanoflagellates are known to be effective filter feeders and are thought to have an advantage over other flagellates when food concentrations are low. In the deep basalt aquifer of the SRPA, choanoflagellates dominate the recovered sequences from the biofilm (70%) and a significant portion of the recovered sequences from the planktonic community (19%, Figure 2 and Table 2 from Strap et al., in review).

A single clone (EUKDBF47) affiliated with the cercozoa made up 16% of the biofilm-derived clones. Cercozoa are small flagellates present in many different environments. It is likely that the organism represented by this clone (EUKDBF47) also grazes on the biofilm-associated bacterial population.

The most common protists as identified by cultural and microscopic studies reported for aquifer environments are flagellates and amoebae. While we did observe flagellates, we did not observe amoebae. It is possible that the amoebae within the SRPA were present in too low a concentration to be detected. ARB analysis indicates representatives of the Acanthamoebidae can be amplified by the NS3/NS8 primer pair.

A single fungal clone (EBF CD5) obtained from the biofilm was found to affiliate with the genus *Emericella.* While a more diverse fungal population was found in the planktonic sample, a planktonic clone (EUKDPK54) was also found to affiliate with *Emericella.* An alignment of the two putative *Emericella* clones showed that they were similar (96%) but not identical.

In the planktonic-derived eukaryote community, 13% of the clones analyzed (11/86) were closely affiliated with *Exophiala* sp., which are dimorphic, black yeasts, members of the order Chaetothyriales, and causative agents of human mycoses. A few notable members of the genus *Exophiala*, such as *Exophiala lecanii-cornii*, *Exophiala jeanselmei*, and *Exophiala oligosperma* have important bioremediation potential due to their ability to degrade a wide range of volatile organic compounds. It is noteworthy that these same species are also pathogenic. The potential pathogenicity and bioremediation potential of the organisms within the SRPA represented by clones EUKDPK66 and EUKDPK64 cannot be determined from their 18S signatures, and thus future work to address this issue would be required.

It is not known whether the natural attenuation of the TCE-contaminated plume of the SRPA occurs within the planktonic population, the biofilm population, or both. Knowledge of the eukaryal populations that coexist with prokaryotic populations is vital

to the future study of natural attenuation in the deep subsurface of fractured basalt aquifers such as the SRPA. The eukaryal populations could potentially impact remediation efforts in a positive manner; for instance, *Exophiala* and other fungi with the capacity for degradation. Also, protists that are able to uptake contaminants may enhance intrinsic bioremediation rates as will protist predation on bacteria that has been demonstrated to increase degradative bacterial growth rates by stimulating bacterial nutrient uptake. By the same token, the eukaryal populations may negatively impact remediation efforts by over-grazing of contaminant-degrading bacteria (and potentially Archaea). The delicate balance of indigenous microbes within contaminated and uncontaminated subsurface environments must be understood.

Knowledge gained by investigations such as that presented here will facilitate interpretation of diversity studies within contaminated aquifer environments.

****

3.0 Specific Accomplishment 3: metagenomic profiling: microarray analysis of an environmental genomic library

Genomic libraries derived from environmental DNA (metagenomic libraries) are useful for characterizing uncultured microorganisms. However, conventional library-screening techniques permit characterization of relatively few environmental clones. In work summarized here, we describe a novel approach for characterization of a metagenomic library by hybridizing the library with DNA from a set of groundwater isolates, reference strains, and communities. A cosmid library derived from a microcosm of groundwater microorganisms was used to construct a microarray (COSMO) containing ~1 kb PCR products amplified from the inserts of 672 cosmids plus a set of 16S ribosomal DNA controls. COSMO was hybridized with Cy5-labeled genomic DNA from each bacterial strain, and the results were compared with the results for a common Cy3-labeled reference DNA sample consisting of a composite of genomic DNA from multiple species. The accuracy of the results was confirmed by the preferential hybridization of each strain to its corresponding rDNA probe. Cosmid clones were identified that hybridized specifically to each of 10 microcosm isolates, and other clones produced positive results with multiple related species, which is indicative of conserved genes. Many clones did not hybridize to any microcosm isolate; however, some of these clones hybridized to community genomic DNA, suggesting that they were derived from microbes that we failed to isolate in pure culture. Based on identification of genes by end sequencing of 17 such clones, DNA could be assigned to functions that have potential ecological importance, including hydrogen oxidation, nitrate reduction, and transposition. Metagenomic profiling offers an effective approach for rapidly characterizing many clones and identifying the clones corresponding to unidentified species of microorganisms.

This work has been published (Sebat, J. L., F. S. Colwell, and R. L. Crawford. 2003. Metagenomic Profiling: Microarray analysis of an environmental genomic library: Comparative genomic hybridization. Appl. Environ. Microbiol. 69:4927-4934) and

readers are referred to this paper for details including tables, figures, and references. This work was supported in part by DOE EMSP award number DE-FG07-02ER63500.

3.1 Approach

3.1.1 Approach overview. Microorganisms contribute significantly to the earth's biological diversity, yet relatively few of the microorganisms present in nature have been cultured and characterized. It is generally accepted that less than 1% of bacteria and fungi present in most habitats have been cultivated for study in pure culture. Although direct analysis of environmental DNA samples by PCR is effective for showing the presence of uncultured microorganisms, biases in primer specificity and amplification of different targets prevent full recognition of microbial diversity. Thus, new approaches to examination of community genomes are needed. The use of large-insert genomic libraries is a powerful approach for isolating DNA sequences from complex mixtures of uncultured microorganisms. Direct cloning of DNA from environmental samples makes it possible to avoid some of the biases of cultivation and PCR. In addition, genomic fragments that are ≥100 kb in length can be obtained, and they provide significant functional and taxonomic information about the organisms from which they were derived. Such metagenomic libraries have been used to identify novel genes from uncultivated species of archaea, bacteria, and viruses that are responsible for significant ecosystem processes and to isolate enzymes that are involved in biosynthesis of novel pharmaceuticals or have other industrial uses.

Given the immense uncultivated and uncharacterized metabolic diversity in the environment, one would need to sequence relatively few bacterial artificial chromosome (BAC) or cosmid clones to discover fundamentally interesting sets of genes. If modern genomic techniques can be used to carry out more comprehensive surveys of metagenomic libraries, our understanding of natural genetic diversity should be greatly enhanced. Screening a genomic library can be done a number of different ways. Typically, screening involves colony hybridization with a probe of interest, which yields information one gene at a time. Bioassays have been developed to screen libraries for genes involved in the production of specific enzymes or natural products; however, this approach relies on the fortuitous expression of heterologous DNA by the library host strain. High-throughput end sequencing of BAC clones has been used to accelerate various single genome projects, and it is currently being used to characterize some environmental DNA libraries. Although the speed and effectiveness of brute-force sequencing are constantly improving, it is not yet practical to assemble a complete bacterial genome from a metagenome. There is still a need for new functional genomic approaches that systematically yield information about many of the elements in a metagenomic library. These new approaches should ideally allow us to identify the organism from which each clone came, to determine some functional characteristics of various clones, and to identify many more novel uncultivated bacteria. We sought to develop a practical approach that would provide a large amount of information about the microbial community from a limited set of clones. The purpose of our approach was to classify many of our cosmids and to identify a few candidates for sequencing rather than to undertake a major sequencing and assembly project. Our method involves hybridization of the library with genomic DNA of various reference strains and bacterial

isolates from the community under study. In addition, DNA derived from as-yet-uncultivated organisms can be identified by hybridization with metagenomic DNA.

3.1.2 Metagenomic profiling. Metagenomic profiling is classification of clones based on hybridization of insert DNA to the genomes of bacterial isolates, reference strains, and environmental DNA. DNA microarray technology has become an important tool for determining the gene contents of entire genomes and measuring the expression of genes. High-density arrays are effective for quantitative detection of genes in complex samples. Thus, microarrays are a promising technique for characterization of genes in environments such as soil and water. However, the use of microarrays has been limited to 16S rRNA markers or a relatively small set of functional genes, and no practical approach has yet been developed to specifically target the unculturable majority of the species in the environment.

We used a microarray platform to screen a metagenomic library with whole microbial genomes and community genomes (Figure 1 from Sebat et al. 2003). The microarray (COSMO) contained ~1 kb PCR products amplified from the inserts of 672 cosmids along with a set of controls (16S ribosomal DNA [rDNA] probes). From an environmental sample (the same sample from which the library was derived), numerous bacterial isolates were obtained. Genomic DNA was purified from each environmental isolate. In addition, metagenomic DNA was purified directly from the mixed population, which was done without cultivation of bacteria. Each test genome was labeled with Cy5-dCTP and probed with COSMO. In order to subtract any signal that may have come from nonspecific hybridization, in each experiment we used two-color hybridization, in which each test genome was compared to a reference sample of common bacterial DNA. The reference DNA consisted of a pooled sample of genomic DNA from 14 species of bacteria (which effectively diluted the strain-specific genes and enriched common sequences). The reference DNA was given a different label (Cy3), and equal amounts of test and reference DNA were combined and hybridized to COSMO. We refer to this approach as comparative genomic hybridization (CGH). CGH was repeated for all environmental isolates, as well as for the metagenomic DNA sample(s). Positive results were determined based on a Cy5/Cy3 ratio greater than 1 (>0 on a $\log_2$ scale). As a result, we obtained a profile for each clone in the metagenomic library (i.e., a graphical representation of its hybridization to one or more species of bacteria). Clones that were specific to a test strain or community hybridized only to those DNA samples. Clones that contained a conserved sequence within their corresponding PCR amplicons hybridized to the genomes of multiple species.

3.2 Materials and Methods

3.2.1 Bacterial strains, media, and culture conditions. To evaluate our approach, we needed a microbial community that could be manipulated in the laboratory. A stable community was developed from an inoculum of biofilm material collected from the Snake River Plain Aquifer in southeastern Idaho. Biofilm was collected from the basalt aquifer by suspending a basket containing 100 mL of ceramic beads in an open borehole at a depth of 73 m adjacent to a zone of high groundwater flow. After 80 days, the basket was retrieved, and the beads were immersed in 100 mL of sterile, phosphate-buffered

saline. Cells were collected by gently vortexing the submerged beads for 15 min. One mL of a cell suspension was used to inoculate triplicate flasks containing 100 ml of minimal succinate medium and 100 mL of glass beads. Minimal succinate medium contained (per L of deionized water) 6.0 g of K_2HPO_4, 3.0 g of KH_2PO_4, 1.0 g of $(NH_4)_2SO_4$, and 4.0 g of succinic acid. The pH of the medium was adjusted to 7.5 with 10 M NaOH. One mL of sterile 1 M $MgSO_4$ and 1.0 ml of sterile 45 mM $CaCl_2$ were added after autoclaving. Microcosms were developed by incubating the flasks without shaking for 1 week at 30°C. After incubation, the cells were suspended by gently vortexing the flask and decanting the medium into two 50-mL polypropylene tubes, and cells were collected by centrifugation and resuspended in 1.0 mL of phosphate-buffered saline.

Approximately 10^{10} cells were used for construction of a cosmid library, and the remaining cells were stored in glycerol at -80°C. To obtain community DNA that was to be used for microarray target samples, a second enrichment of the microcosm was carried out under the same conditions by using a 0.5% inoculum of frozen stock. After incubation, two different fractions of cells were collected (pellicle and planktonic). First, the pellicle was removed with a sterile spatula and placed in a 50-mL polypropylene tube, and then the remaining cells were suspended by gently vortexing the flask, decanting the medium into two 50-mL polypropylene tubes, and collecting the cells by centrifugation. DNA from the samples described above was probed with COSMO in order to identify clones corresponding to microbial strains that were enriched in the mixed culture but that we failed to isolate in pure culture. We used the unisolated fraction of the microcosm to represent the uncultured microorganisms in the environment.

From the original microcosm and all subsequent enrichments, bacterial species were isolated by plating serial dilutions of liquid cultures onto plates containing tryptic soy agar. Bacterial strains were identified by selecting colonies with unique morphology that appeared during a 7-d incubation at 30°C. Ten different strains were obtained (see Figure 4 from Sebat et al. 2003) along with the *Pseudomonas aeruginosa* and *Staphylococcus aureus* reference strains. To determine the identity of each isolate, the 16S rRNA gene was amplified by PCR from a genomic DNA template using eubacterial primers 27F and 907R. 16S amplicons were sequenced using the 27F primer.

3.2.2 Library construction. A cosmid library was constructed from the genomic DNA of the original mixed bacterial enrichment. Bacterial cells were embedded in agarose, and genomic DNA was purified by agarose-embedded cell lysis, followed by partial digestion of the agarose plugs with *Sau*3AI. The metagenomic library was constructed by using a SuperCos I cosmid kit (Stratagene Corp., La Jolla, CA) according to the manufacturer's protocols. Clones were picked randomly into 96-well plates containing Luria-Bertani medium supplemented with 100 mg L^{-1} of ampicillin and 0.1 volume of 10 x Hogness buffer (40% glycerol, 36 mM K_2HPO_4, 13 mM KH_2PO_4, 20 mM trisodium citrate, and 10 mM $MgSO_4$ in deionized water). After overnight incubation, the library was stored at -80°C. Plasmids were purified by using a REAL prep 96 kit (Qiagen) and a BioRobot 3000 (Qiagen) liquid-handling system. When we examined *Xho*I restriction digests of 10 clones by agarose gel electrophoresis, we observed no duplicate clones, and we determined that the average insert size of the cosmids was ~40 kb. The results presented below verify that COSMO represented much of the microcosm's diversity. We did not

attempt to characterize the species represented in the cosmid library prior to fabrication of COSMO. The task of amplifying 16S rRNA genes from a pool of cosmids was confounded by the presence of contaminating *Escherichia coli* genomic DNA in the plasmid preparation.

3.2.3 Array fabrication. All microarray experiments were performed with COSMO, a DNA microarray containing end fragments of 672 cosmids selected randomly from the metagenomic library plus a set of reference genes (16S rDNA markers from several microcosm isolates [Figure 2 from Sebat et al. 2003]). Cosmid end fragments were produced by a thermal asymmetric interlaced PCR method. Briefly, the thermal asymmetric interlaced PCR method involved sequential cycles of linear amplification of insert DNA from the T7 end of the vector with nested primers, followed by exponential amplification of the specific product by random priming, which resulted in PCR products that were 200-2,000 bp long. High-throughput PCR was performed with an MBS 384S thermocycler system (ThermoHybaid). 16S rRNA gene controls were also amplified by PCR. The quality and quantity of DNA were confirmed by agarose gel electrophoresis. PCR products were purified by using 384-well filter plates (Millipore) and were resuspended in 15 μL of 1x Spotting Solution Plus (Telechem Intl., Sunnyvale, CA) to obtain a final DNA concentration of 100-200 ng μL^{-1}. Each sample was spotted in duplicate on SuperAmine slides (Telechem) by using a Microgrid arrayer (BioRobotics, Cambridge, UK). Slide cross-linking, washing, and blocking steps were carried out according to the manufacturer's protocols (http://arrayit.com/PDF/Super_Microarray_Substrates.pdf).

3.2.4 DNA preparation, labeling, and hybridization. Genomic target DNA was purified from bacterial isolates and mixed cultures. The reference sample was prepared by mixing equal amounts of genomic DNA from the 10 species of bacterial isolates that were used in this study and from the reference organisms *E. coli*, *P. aeruginosa*, *S. aureus*, and *Bacillus subtilis*. Fluorescently labeled target DNA was prepared by established techniques. Briefly, 2 μg of target DNA was digested completely with *Msp*I and purified by ethanol precipitation. Prior to labeling, 10 ng of *salT* (a rice gene used as an internal standard) was added to the sample. Target DNA was labeled by incorporation of Cy5-dCTP (for test DNA) or Cy3-dCTP (for reference DNA) (Amersham Pharmacia, Pittsburgh, PA) by random primer synthesis (BioPrime labeling kit; Invitrogen). Labeled target DNA was purified with CHROMA SPIN + TE-30 gel filtration columns (Clontech Laboratories, Mountain View, CA). Test DNA and reference DNA were combined with 40 μg of human Cot-1 DNA and 100 μg of salmon sperm genomic DNA and reduced to a volume of 5 μL by using Microcon YM-30 concentrators (Millipore); 20 μL of 1.25 x Unihyb hybridization buffer (Telechem) was added to the target DNA mixture, and the target was preannealed to blocking DNA by boiling the preparation for 1.5 min, followed by 30 min of incubation at 37°C. Twenty-five μL of probe was used per slide. Hybridization was performed for 8 h at 65°C. Post hybridization washing was performed according to the slide manufacturer's protocol.

3.2.6 Data analysis. Arrays were scanned with an Axon 4000 scanner, and fluorescence measurements were obtained by using Genepix Pro 3.0 software (Axon Instruments,

Sunnyvale, CA). Data sets were filtered for spots with a signal-to-noise ratio greater than 3.0 using Microsoft Excel. Results are reported below as $\log_2$ of the Cy5/Cy3 ratio. Data analysis and graphical display were conducted using Expression NTI (Informax Inc.) in the following manner: 1) all clones that failed to produce a positive result (log2, >0) in duplicate experiments with at least one target genome were removed from the data set, 2) clones were clustered by complete linkage of genes using the correlation coefficient, and 3) experiments were sorted according to the phylogenetic relationships of test strains as determined by Clustal W alignment of 16S rRNA sequences.

3.2.6 Analysis of cosmid end sequences. The inserts of select clones were sequenced from both ends of the multiple cloning site by using the T3 and T7 primers. The plasmid template was purified from 500-mL cultures of *E. coli* by using a large construct kit (Qiagen). A sequencing reaction mixture (total volume, 20 µl) was prepared by combining 1 µg of plasmid DNA with 0.32 µL of a 10 µM primer T3 or T7 stock solution in Tris-EDTA buffer and 8.0 µL of a Big Dye mixture. PCR was performed with a PTC-100 thermocycler (MJ Research, Ramsey, MN) for 80 cycles (94°C for 30 s, 47°C for 15 s, and 60°C for 4 min). Reaction mixtures were purified with DTR gel filtration columns (Edge Biosystems, Gaithersburg, MD). Nucleotide sequences (average size of 450 bp) were determined with an ABI 3100 DNA sequencer. The potential functions and phylogenetic affinities of cosmid end sequences were determined by performing a nucleotide and translated-protein search of GenBank using the Basic Local Alignment Search Tool (BLAST). The potential function of a given sequence was determined by examination of all homologous sequences with expect values of <1e-2. A particular function was assigned if 1) a consensus was apparent among the best hits and 2) there was no disagreement between the consensus of the nucleotide results and the translated protein results. We also listed the species corresponding to the nucleotide best hit; when no significant nucleotide matches were observed; we listed the species corresponding to the protein best hit (Table 1 from Sebat et al. 2003).

3.3 Results

3.3.1 Data validation. A single self-versus-self hybridization was performed by hybridizing COSMO with two samples of *Acidovorax* B genomic DNA that had been prepared and labeled separately, one with Cy5 and the other with Cy3. Seventy-six probes corresponding to 43 different cosmids passed our filtering criteria (see Section 3.2 Materials and Methods). The values for the Cy5/Cy3 ratio ranged from 0.77 to 1.19 (mean, 1.00; standard deviation, 0.12). Based on this experiment, the performance of different probes and dyes should have caused less than a 1.3-fold deviation from the mean. The sensitivity and specificity of our microarray analysis were evident from the results obtained for controls and reference genes (Figure 2 from Sebat et al. 2003) in our CGH experiments with individual genomes (see below). 16S rRNA genes from several microcosm isolates were obtained by PCR and included in COSMO as controls. Each rRNA gene served as a positive control for its corresponding species and as a negative control for distantly related species. Additional negative controls are shown last in the figures. The results indicated that the hybridization conditions used allowed identification of strain-specific genes. In all but one case, genomic DNA of each strain hybridized preferentially to its corresponding rRNA probe (yielding the highest ratio). In most cases,

some hybridization to a related strain was observed, but the level of hybridization was lower. In the case of *Ultramicrobacterium* 3, genomic DNA hybridized nearly equally to the *Ultramicrobacterium* 3 and *Acidovorax* B rRNA probes. No positive results were obtained with the *Bacillus* sp., *Shigella* sp., and *Staphylococcus* sp. negative controls or with the cosmid vector.

3.3.2 Hybridization with individual genomes. Various reference strains and microcosm isolates were used individually as target DNA in experiments to locate clones related to these organisms. The clustered microarray results for various microcosm isolates and reference strains revealed distinct classes of DNA that corresponded to individual strains or groups of bacteria (Figure 3 from Sebat et al. 2003). Numerous strain-specific clones were apparent. The data also revealed examples in which clones hybridized to multiple related organisms (indicative of conserved genes). Based on these patterns, we classified some clones as members a particular species, genus, or branch (Figure 3 from Sebat et al. 2003).

At the bottom of Figure 3 (see Sebat et al. 2003), some results appear to be scrambled (i.e., distinct patterns are not easily distinguished). The profiles observed for this group of clones are not consistent with profiles of strain-specific and conserved clones. In general, these clones appeared to cross-hybridize between species to a greater extent. A variety of different patterns were observed, and for the sake of simplicity, they were not labeled.

3.3.3 Classifying uncultured DNA. In the experiment described above (Figure 3 from Sebat et al. 2003), 156 probes (clones) passed our filtering process. The remaining 524 clones failed to produce a positive $\log_2$ ratio with any test strain or failed to produce any significant signal. Some of these clones may have corresponded to organisms present in the microcosm that we failed to isolate in pure culture. We considered the unisolated organisms in our microcosm to be analogous to uncultured microbes in the environment. To identify cosmids derived from such organisms, we performed a similar experiment using genomic DNA extracted directly from the mixed bacterial population as the test DNA (the same reference DNA was used). The results of two metagenomic CGH experiments were added to the data set prior to clustering (Figure 4 from Sebat et al. 2003). The results identified a number of clones that were present in the community and not in our catalog of isolates (Figure 4A from Sebat et al. 2003). Such clones were classified as uncultured.

A single experiment, such as the one described above, yielded a spectrum of ratios. It is likely that the clones that yielded a $\log_2$ ratio of 1 corresponded to a different organism than the clones that yielded a $\log_2$ ratio of 6. The uncultured class could be separated into subgroups if there were clear differences in the abundance of different genes in the community, but it was not obvious where to draw the line between one organism and another.

In an attempt to resolve the uncultivated community in slightly better detail, the microcosm cells were fractionated into two types, pellicle cells and free-swimming cells, which were analyzed separately. All uncultured clones (Figure 4A from Sebat et al. 2003)

were found to have the same even distribution between planktonic and pellicle cells; therefore, we were not able to determine if these clones corresponded to multiple species. However, we observed that clones corresponding to some of the cultivable species had distinct distribution patterns. *Acidivorax* I was apparently distributed throughout the microcosm (Figure 4B from Sebat et al. 2003), *Acidivorax* B was present primarily in the pellicle (Fig. 4C), and *Caulobacter* 6 was not abundant in either fraction (Figure 4D). These clusters were labeled and presented as examples of clones from different species that, in principle, could be distinguished based solely on community analysis.

3.3.4 Sequence analysis of uncultivated DNA. In order to identify functional characteristics of uncultured microorganisms from the microcosm, insert DNA from each of 17 random clones from cluster A (Figure 4A from Sebat et al. 2003) was sequenced from the T7 and T3 primer sites that flanked the insert. A nucleotide and translated-protein search of GenBank was performed with each cosmid end sequence (Table 1 from Sebat et al. 2003). Among 34 sequences, 18 functional genes were identified, 10 hits were obtained with genes of unknown function, and six sequences yielded no significant result. The great majority of the sequences were found to be significantly similar to genes from members of the *Proteobacteria*, including seven genes from *Ralstonia* sp. Some of the sequences could be assigned to functions having ecological importance, including a putative [NiFe] hydrogenase, nitrate reduction, and several transposases. Four different insertion sequence elements (ISs) were observed in five clones, 2H6, 6A12, 6A7, 6D10, and 7A12 (sequences from 6A7 and 6D10 were from different positions of the same gene). All ISs that we identified occurred precisely in the T7 end fragment.

3.4 Discussion

The experiments described above illustrate a useful technique for rapidly classifying DNA from a metagenomic library and, in the process, identifying conserved and divergent genome fragments from a particular community, genus, strain, or species and DNA corresponding to strains that have not been isolated in pure culture.

Hybridizing COSMO with single genomes and cluster analysis of the microarray results were effective for visualizing groups of clones that have unique patterns of hybridization to different bacterial genomes (Figure 3 from Sebat et al. 2003). We identified clones that hybridized to a single strain, as well as clones that hybridized to related species. In this manner, we classified clones as strain-specific or specific to a broader phylogenetic group. Many of the clones in the scrambled section of Figure 3 (Sebat et al. 2003) have much broader specificity, and some cross-hybridize only between two distantly related strains. The latter finding may be of interest to researchers who wish to identify genes that have been transferred horizontally between species. However, before such claims are made based on hybridization between more divergent species, the correlation between homology and signal intensity must be examined more carefully. Identification of clones corresponding to cultivable strains is also useful for eliminating clones that do not need to be examined in a culture-independent manner. This may in fact be the primary objective, in which case it would not be necessary to probe genomes one at a time, as we did. The cultivable genomes could be combined into pools in order to accelerate the process.

The eventual identification of clones corresponding to uncultivated microorganisms is the most valuable aspect of this approach. We identified clones that hybridized uniquely to total community DNA. These clones corresponded to one or more species of bacteria that were enriched in mixed cultures but that we were unable to obtain by isolation on tryptic soy agar. End sequencing of 17 cosmids resulted in identification of numerous sequences that had nucleotide similarity to the genes of *Ralstonia* spp. In addition, we identified a putative hydrogenase and genes involved in reduction of nitrate. Some species of bacteria, including *Ralstonia* spp., have the ability to couple hydrogen oxidation to nitrate reduction. The putative hydrogenase gene and the *norE* homologue were not found in the same clone, and we could not confirm that they are linked in any way. Obtaining such evidence would require isolating both sets of genes on the same fragment of DNA. However, knowledge of the metabolic genes present in the uncultured population provides information that should be important for selectively culturing such organisms if they are indeed present.

Among the clones that hybridized only to community DNA, a remarkably high frequency of ISs was observed. We do not believe that this is a unique characteristic of the uncultivated species. ISs occur in a wide range of genera. A variety of ISs can occur in a single genome, and a single type of IS may have a copy number of >20. We believe that the five occurrences of four different IS-like elements precisely in a clone's T7 end fragment (the probe actually printed on the array) may have reflected a greater sensitivity of our method for high-copy-number sequences.

Of course, in a mixed population, our method detects the most abundant genes. A common limitation of microarray studies is sensitivity. The detection limit for microarray analysis of soil samples is on the order of ~50 ng of a single bacterial genome; therefore, in a typical experiment one would detect only organisms that constitute at least 1/40 of the population.

We hope to further develop applications of the COSMO microarray for environmental samples. We demonstrate here that in addition to simple identification of an uncultured subset of clones, additional classes of organisms can be defined by comparing different environmental samples from the same habitat (Figures 4B to D from Sebat et al. 2003). Clones that have a unique distribution in the two fractions may constitute a separate phylogenetic class. By comparing the metagenomic profiles of a field site to a map of metabolic activities or microbial species, clones whose distribution correlates with a biological process may be identified. In this manner, ecologically important genes for which there is not a specific probe or assay may be identified.

Systematic organization of a metagenomic library is essential for performing a more comprehensive study of a microbial community. We must continue to utilize modern techniques for genome analysis and adapt them to the study of complex mixed genomes, keeping in mind that the purpose of a genome-wide study is to accelerate the discovery of genes that are important for specific processes.

A study of the microbial metagenome need not seek truly comprehensive knowledge concerning all microbial genomes if it is possible to first negotiate the genomic landscape of a given site and find what is relevant and interesting. A practical approach to metagenomics is 1) to quickly identify familiar genes, 2) to identify the unknowns, and 3) to attempt to classify them based on knowledge such as where certain genes are present and what apparent linkages there are between different genes. Once a set of clones that is linked to a biological process is identified, the specific genes involved may be identified from the complete DNA sequences of the clones. Metagenomic profiling is an appropriate technique for these tasks.

We believe that microarray studies in combination with DNA sequence analysis will be important tools for enhancing our understanding of Earth's microbial diversity.

****

ATTACHMENTS (IN PDF OR WORD FORMAT)

1. Sebat, J. L., F. S. Colwell, and R. L. Crawford. 2003. Metagenomic Profiling: Microarray analysis of an environmental genomic library: Comparative genomic hybridization. *Appl. Environ. Microbiol.* 69:4927-4934.
2. Erwin, D. P., I. K. Erickson, M. E. Delwiche, F. S. Colwell, J. L. Strap, and R. L. Crawford. 2005. Diversity of oxygenase genes from methane- and ammonia-oxidizing bacteria in the Eastern Snake River Plain aquifer. *Appl. Environ. Microbiol.* 71:2016-2025.
3. Strap, J. L., F. S. Colwell and R. L. Crawford. Eukaryotic diversity in planktonic and biofilm populations of the Snake River Plain Aquifer. *Microbial Ecology.* In review.

Eukaryotic Diversity in Planktonic and Biofilm Populations of the Snake River Plain Aquifer

Janice L. Strap[1]*, Frederick S. Colwell[2], and Ronald L. Crawford[1]

[1] University of Idaho, Environmental Biotechnology Institute, Moscow, Idaho, 83844-1052

[2]Idaho National Laboratory, Biological Sciences, Idaho Falls, Idaho, 83415-2203

[*] Corresponding author. Mailing address: University of Idaho, Environmental Biotechnology Institute, Food Research Center 103, Moscow, Idaho 83844-1052, Phone: (208) 885-5893; Fax: (208) 885-5741, E-mail: jstrap@uidaho.edu

ABSTRACT

Eukaryotic microorganisms are an important component of microbial communities in aquifers, yet phylogenetic information with respect to their diversity within this habitat is scarce. We characterized by molecular methods, the eukaryotic microbial diversity of planktonic and biofilm communities within the oxic, fractured basalt Snake River Plain Aquifer (SRPA). Planktonic microorganisms were collected by filtration of pumped groundwater and biofilms were obtained by colonization of basalt chips housed within substrate columns suspended in the aquifer. 18S rDNA sequences were placed into operational taxonomic unit (OTU) groups with ≥99% 18S rDNA similarity. The planktonic community was composed of fungi (49%), stramenopiles (32%), and choanoflagellida (19%), while the biofilm community was comprised of choanoflagellida (70%), cercozoa (16%), and fungi (14%). Three novel fungal clades, five novel stramenopile clades, and seven novel choanoflagellate sequences were identified. The planktonic community harbored greater diversity than the corresponding biofilm as indicated by the reciprocal of Simpson's index (1/D) and the Shannon-Weiner index. Non-parametric estimation of the species richness by Chao1 revealed that fewer species were expected in the biofilm (5 species) population than in the planktonic community (30 species). The ∫-LIBSHUFF program provided evidence that the planktonic-derived OTUs were significantly ($p = 0.0001$) different from those of the biofilm. This was confirmed by F_{ST} analysis ($F_{ST} = 0.08497$, $p<0.0001$). While planktonic protists have been previously described in aquifers, to our knowledge, this is the first report of planktonic and biofilm-associated eukarya characterized from an aquifer by molecular methods.

INTRODUCTION

Molecular methods have been extensively used to investigate prokaryotic diversity in a wide range of environments, however molecular investigations of eukaryotic populations are more recent and have not yet been as extensive. The reason for this is in part due to a well-established ultra-structure and morphology-based eukaryote taxonomy. However, morphological classification can be complicated by the sheer number, diversity and size range of many species (10) within an environmental sample. The molecular studies of eukaryotes which have been conducted have focused on soils (43), desert rock varnish (41), rivers (3), solar salterns (11), marine environments (18, 23, 47, 55, 58, 66, 73), anoxic sediments (19), and acid mine drainage (5). There have been relatively few prior reports of eukaryotic diversity within aquifers with much of the work consisting of visual/microscopic and/or culture-based evaluations (25, 38-40, 48, 50). Microeukaryotes have a direct impact on the ecology of the aquifer due to their important role in the food web of this environment. For example, the consumption of bacteria and archaea by protists affect the abundance of these organisms and thereby play a principal role in nutrient cycling (4, 75).

Studies of biofilms have historically focused on prokaryotes and therefore, knowledge of substrate-associated eukaryotes is sparse. It is likely that microeukaryotes such as protists influence the architecture of biofilm by their own presence in addition to their feeding on substrate associated bacteria and archaea (7, 56). Population dynamics of protozoa in biofilms has been investigated *in vitro* (35, 56, 79), but a molecular characterization of microeukaryotes of biofilms sampled from deep, oxic aquifers to our knowledge has not been conducted.

58 The rather limited number of currently available 18S rRNA sequences has a
59 negative impact on the ability to design appropriate primers for the culture-independent
60 assessment of microbial diversity. One way this limitation can be overcome is to add to
61 the inventory of microeukaryotic phylotypes.
62 Investigations into the diversity of microeukaryotes within deep aquifers will give
63 insight into key relationships between species richness and organic contamination (62).
64 This is particularly important in areas such as the Test Area North (TAN) site of the
65 Idaho National Laboratory (INL) in southeast Idaho where a trichloroethylene (TCE)
66 plume contaminates the Snake River Plain Aquifer (SRPA) that underlies the site. The
67 concentration of dissolved TCE at the site is decreasing by natural attenuation and this
68 has positive, economic impact for remediation of the site (72). The participation of
69 eukarya in intrinsic bioremediation should not be overlooked. To fully understand the
70 process of natural attenuation, knowledge of the indigenous eukarya in pristine regions is
71 required.
72 Here we report the planktonic and biofilm microeukaryotic diversity found within
73 a pristine region of the oxic, fractured basalt SRPA. It was demonstrated that the
74 planktonic community was more diverse than the biofilm community. To our knowledge
75 this is the first report of an assessment of eukaryotic diversity of an aquifer environment
76 by molecular methods.
77

78 **MATERIALS AND METHODS**

79 **Sampling.** Two wells in a pristine region of the Snake River Plain Aquifer near the Test
80 Area North (TAN) site at the Idaho National Laboratory were used for sampling. The

81 planktonic community was sampled from well ANP-9 by concentrating the cells present
82 in 13,300 L of aquifer water to approximately 3 L by employing a KrosFlo Pilot hollow-
83 fiber, tangential flow filtration system with a DynaFibre microporous membrane (0.2 μm
84 pore size and 3.9 m^2 surface area; Spectrum Labs, Rancho Dominguiz, Calif.). The well
85 was flushed with 3 well volumes (1 volume equal to the standing water in the well) to
86 remove built-up sediments and debris from the casing prior to filtration. The concentrate
87 was collected by reversing the pump motor and directing the retentate into sterile 1 L
88 containers which were subsequently stored at -80ºC and thawed at room temperature
89 prior to use (26). The biofilm community was sampled from well ANP-10 by incubating
90 sterile basalt substrates (4 to 7 mm sieve size) housed in 76 cm long nylon mesh tubes
91 (ID, 2 cm; mesh size 0.25 cm; InterNet, Inc., Minneapolis, Minn.) in the aquifer for six
92 months at a depth of 80 m (26). The wells chosen for the study are in close proximity
93 (ca. 200 m) to each other. Previous analysis indicated comparable water chemistry
94 between the two wells and that the values are typical of the SRPA. For example a
95 temperature of 14.1 to 14.5ºC, pH of 7.2 to 7.9, dissolved oxygen level of 7.5 to 8.4 mg
96 L^{-1}, dissolved organic carbon level of 1.5 to 2.6 mg L^{-1}, and methane concentration of 72
97 nmol L^{-1} have been reported (60).
98
99 **Total DNA Isolation.** Removal of biofilm biomass from the basalt substrates was
100 achieved by filling four 50-mL polypropylene tubes with 40 g of the basalt substrates and
101 incubating at 200 rpm at 60ºC for 30 min in a horizontal position in a rotating incubator
102 (Innova 4300; New Brunswick Scientific Co., Inc., Edison, NJ).

103 Planktonic cells were recovered by filtration of the concentrated material through
104 a 25 mm (0.2 μm pore size) Nuclepore polycarbonate membrane filter (Whatman, Inc.,
105 Clifton, NJ).
106 DNA from the planktonic and biofilm communities were isolated as described in
107 (26) using the UltraClean™ Soil DNA Kit Mega Prep (MoBio Laboratories, Inc., Solana
108 Beach, Calif.) following the manufacturer's instructions. This method has been shown to
109 be effective for extraction of DNA from both prokaryotes and eukaryotes (9, 16, 22, 29,
110 31, 61). Two independent extractions from each community were pooled for use as
111 template. The resulting DNA was stored at -20ºC.
112
113 **Clone Library Construction.** Eukaryote 18S rRNA genes were amplified from total
114 DNA template using NS3 (nu-SSU-0505-5'; 5'-GCAAGTCTGGTGCCAGCAGCC-3') and
115 NS8 (nu-SSU-1769-3'; 5'-TCCGCAGGTTCACCTACGGA-3') primers (76). These
116 primers were previously described for amplification of fungal rDNA but analysis of the
117 primer sequences by BLAST (1) and by ARB (49) revealed this primer pair had a broad
118 specificity including Metazoa, Viridiplantae, Fungi, Ichthyosporea, Acanthamoebidae,
119 Haptophyceae, Cryptophyta, Rhodophyta, Stramenopiles, Cercozoa, Alveolata,
120 Polycystinea, Diplomonadida, Apusomonadidae, Glaucocystophyceae, Choanoflagellida,
121 Mesomycetozoa, and Epsilon-proteobacteria. Each PCR mixture (50 μL) contained 1X
122 PCR buffer (Sigma-Aldrich, St. Louis, Mo.), 2.5 mM $MgCl_2$, 200 μM (each)
123 deoxyribonucleoside triphosphate (dTTP, dATP, dGTP, and dCTP), 1.0 μM (each)
124 primer, and 20 to 50 ng template DNA and JumpStart™ *Taq* DNA polymerase (0.05 U;
125 Sigma-Aldrich). PCR amplification was performed on an MBS Thermal Cycler

126 (ThermoHybaid) under the following conditions: initial denaturation of the template
127 DNA at 95ºC for 5 min, 25 cycles of 95ºC for 30 s, 55ºC for 30 s, and 72ºC for 60 s, and
128 a final elongation step at 72ºC for 10 min. The PCR products were visualized by
129 electrophoresis on 1% (w/v) agarose gels. Positive controls consisted of reactions
130 containing DNA from *Candida albicans* ATCC 90027 and *Aspergillus nidulans.*
131 Negative controls consisted of PCR reactions containing no template DNA addition,
132 *Archaeoglobus fulgidus* ATCC 49558, *Halobacterium salinarum* ATCC 19700 or
133 *Escherichia coli* DNA as template. To minimize PCR bias, four independent PCR
134 reactions were pooled. The combined reaction products were purified with MO BIO
135 Ultraclean™ PCR Clean-Up™ kit (Mo Bio Laboratories Inc., Carlsbad, Calif.) according
136 to the manufacturer's instructions. The amplicons were cloned using the TOPO TA
137 Cloning® Kit for Sequencing (version k) (Invitrogen Corp., Carlsbad, Calif.). Clones
138 were grown in 96 well format plates and plasmid DNA was purified using the Montage
139 Plasmid Miniprep$_{96}$ Kit (Millipore, Billerica, Ma).
140
141 **Restriction fragment length polymorphism (RFLP) Analysis.** RFLP patterns were
142 produced for each clone by digestion of 1 µg plasmid DNA with 5 U of *Hae*III
143 (Invitrogen) in a final reaction volume of 50 µL. Reactions were incubated at 37ºC for 2
144 hr followed by enzyme inactivation at 65ºC for 10 min. Fragmentation patterns were
145 analyzed by agarose gel electrophoresis on 3.5% NuSieve 3:1 agarose (Cambrex Bio
146 Science, Baltimore, MD). Gels were stained with ethidium bromide and visualized under
147 UV transillumination. Kodak 1D Image analysis software (Kodak Eastman, New Haven,
148 CT) was used to facilitate grouping of clones.

Sequence and Phylogenetic Analysis. One representative from each RFLP group was sequenced in both directions using T3 and T7 primers. Nucleotide sequence determinations of each representative clone was performed on an ABI Prism Model 3730 (Applied Biosystems, Inc., Foster City, Ca) using Big Dye version 3 chemistry at the Laboratory for Biotechnology and Bioanalysis (School of Molecular Biosciences, Washington State University, Pullman, WA). Contiguous sequences were obtained using ContigExpress (Vector NTI suite version 9). Prior to comparative sequence analysis, vector sequences flanking the 18S rDNA inserts were manually removed. Sequences were analyzed for chimeras using the Ribosomal Database Project II CHECK_CHIMERA program (51, 52) and Bellerophon (32). Three putative chimeric planktonic sequences and one putative chimeric biofilm sequence were detected and removed from analysis. Nucleotide-nucleotide BLAST (Basic Local Alignment Search Tool) (1) was used to search GenBank for nearest relative sequences. BLAST results, representative eukarya sequences and RFLP representative clones were aligned using ClustalX (74) and MUSCLE (24). Very variable regions of the alignment were removed with Gblocks (12) using parameters optimized for rDNA alignments (minimum length of a block, 5; allowing gaps in half positions), leaving 1307 informative positions. Edited alignments were evaluated with the Maximum Likelihood (ML) method using PAUP* (version 4.0b10; Sinauer Associates, Inc., Sutherland, Ma) under the TrNef+G model of evolution which was determined to be the model best describing the data as determined by the DT-ModSel program (57). PAUP* was run on a cluster computer that has 130 dual Xeon processors and on a Silicon Graphics Octane 2 workstation. Confidence

172 estimates for the nodes within the phylogenetic trees were performed by bootstrap
173 analysis (1000 replicates). Trees were visualized with TreeView software version 1.6.6
174 (64). Maximum Parsimony (MP) bootstrap values were computed with PAUP by using a
175 heuristic search method with a tree bisection-reconnection branch-swapping option with
176 random taxon addition.
177
178 **Rarefaction, clone coverage and species richness estimation.** Diversity coverage by
179 the clone libraries was analyzed with DOTUR (67). Clone coverage was estimated as
180 described by Mullins et al. (59). The percent coverage (C) of the clone libraries was
181 calculated according to the equation $C = [1-n_1/N] \times 100$ where n_1 is the number of unique
182 clones and N is the total number of clones analyzed (59). The eukaryotic species richness
183 was calculated for both the planktonic and biofilm communities using the nonparametric
184 estimators ACE (abundance-based coverage estimator) and Chao1 (13, 14, 33) using
185 DOTUR (67).
186
187 **Statistical comparison of coverage.** The planktonic and biofilm 18S rDNA libraries
188 were compared using the ∫-LIBSHUFF (68) (http://www.plantpath.wisc.edu/fac/joh/S-
189 libshuff.html) computer program which used the coverage formula of Good (30) to
190 generate homologous and heterologous coverage curves from the 18S rDNA clone
191 libraries. Sequences were randomly shuffled 999 times between samples prior to
192 calculating the distance between the curves using the integral form of the Cramér-von
193 Mises test statistic (68). The DNADIST program of PHYLIP
194 (http://evolution.genetics.washington.edu/phylip.html) using the Jukes-Cantor model for

195 nucleotide substitution was used to generate the distance matrix analyzed by ∫-
196 LIBSHUFF. The ∫-LIBSHUFF program provides a statistical test for the null hypothesis
197 that two 18S rRNA gene libraries are samples of the same community.
198
199 **Comparison of phylogenetic diversity.** The phylogenetic diversity between the
200 planktonic and biofilm community was compared using *F* statistics (F_{ST}) (53) calculated
201 using the ARLEQUIN program (69). The F_{ST} test was used to compare the genetic
202 diversity within each community to the total genetic diversity of the communities
203 combined, using the equation $F_{ST} = (\theta_T - \theta_W)/\theta_T$, where θ_T is the genetic diversity for all
204 samples and θ_w is the genetic diversity in each community (53). Statistical significance
205 of F_{ST} was evaluated by randomly assigning sequences to populations and calculating F_{ST}
206 for 1,000 permutations. An $F_{ST} \approx 0$ indicates that genetic diversity overlaps between the
207 communities being compared (53). As the difference between the populations increases
208 the F_{ST} value also increases.
209
210 **Diversity Indices**. Diversity indices utilized to describe and compare the biofilm and
211 planktonic libraries for eukaryotic diversity were determined as described by Martin (53).
212 The reciprocal of Simpson's index (1/D) and the Shannon-Weiner index were used as
213 general diversity measures (78) and were calculated in DOTUR (67).
214
215 **Nucleotide Sequence Accession Numbers.** The sequences of the cloned inserts were
216 deposited in GenBank under accession numbers DQ104580 – DQ104608.
217

RESULTS

Planktonic community structure. Total DNA extracted from groundwater at well ANP-9 at the TAN site was used to construct a clone library of partial eukaryotic 18S rDNA genes. Initially, 89 planktonic clones were screened by RFLP analysis and grouped according to RFLP type. Three putative chimeras were removed from further analysis. Twenty-four unique phylotypes were obtained from the remaining 86 non-chimeric clones analyzed. The phylogenetic tree in Fig. 1 shows the relationship among the 24 planktonic phylotypes. The eukaryotic planktonic community was comprised mainly of fungi (49%) with the choanoflagellates and stramenopiles accounting for 19% and 32% of the clones analyzed respectively (Fig. 2). Two novel planktonic clades belonging to the fungi (clade FII, EPK BH1 and EPK B G4; clade FIII, EUKDPK65, EPK B B11, EPK B G7 and EUKDPK41) and one belonging to the choanoflagellida (EUKDPK38, EPK B E10, EPK B C12, EPK B C11 and EPK B C4) were observed (Fig. 1). Four novel clades within the stramenopiles were also identified within the planktonic community: clade SI, represented by EPK B G5 and EPK B F7; clade SII, represented by EUKDPK45; clade SIII, represented by EUKDPK52, EUKDPK44, EUKDPK69; clade SIV, represented by EPK B A10 and EPK B C3; and clade SV, represented by EPK B C10. Clade SIII was most closely affiliated with an uncultured clone obtained from Spain's 'River of Fire' (3) while the other novel stramenopile clades did not affiliate with any known sequences.

Biofilm community structure. Total DNA recovered from colonized basalt chips after submersion for 6 months in well ANP-10 was used to construct a clone library of partial

241 eukaryotic 18S rDNA genes. Eighty-one non-chimeric clones were screened by RFLP
242 analysis. Of these clones, it was found that 44 sequences (representing 35 distinct RFLP
243 groups) were determined to be archaeal in origin and were subsequently excluded from
244 this study. From the 37 remaining eukaryal biofilm clones, 5 distinct RFLP patterns were
245 observed. It is interesting to note that the NS3/NS8 primer set did not yield an amplicon
246 when the archaea *Archaeoglobus fulgidus* or *Halobacterium salinarum* DNA was used as
247 template. Furthermore, no archaeal sequences were amplified with these primers when
248 the total DNA from the planktonic community was used as template. The phylogenetic
249 tree in Fig. 1 shows the relationship among the 5 biofilm phylotypes. A novel clade
250 belonging to the choanoflagellida (EBF C D10 and EBF C E3) was observed. The
251 eukaryotic biofilm was composed mainly of choanoflagellates (70%, Fig. 2), fungi (14%,
252 Fig. 2) and cercozoa (16%, Fig. 2).
253
254 **Rarefaction analysis.** Rarefaction analysis of the RFLP results obtained for both the
255 planktonic- and biofilm-derived communities was performed to estimate the extent to
256 which the eukaryal diversity of the two communities could be described by screening 86
257 and 37 non-chimeric clones respectively. Rarefaction analysis results for the two clone
258 libraries revealed that the number of planktonic clones analyzed was insufficient to
259 describe the total extent of eukaryotic diversity since no obvious plateau was observed
260 (data not shown). Thus, further analysis of 18S rDNA clones for the planktonic
261 community would likely have revealed additional diversity. In contrast, the rarefaction
262 curve for the biofilm clones did reach saturation and indicates that the diversity within the
263 biofilm community had been adequately sampled (data not shown). Percent coverage was

264 determined to be 72% and 87% for the planktonic and biofilm libraries, respectively. The
265 higher coverage of the biofilm is consistent with the lower diversity observed in this
266 library.
267
268 **∫-LIBSHUFF comparisons of the planktonic and biofilm communities.** To determine
269 whether the planktonic and biofilm 18S rDNA libraries were samples of the same
270 communities, ∫-LIBSHUFF analysis was performed. The ∫-LIBSHUFF (68) comparisons
271 of the planktonic and biofilm communities suggested that the two communities were
272 derived from populations of different composition (data not shown). ∫-LIBSHUFF
273 analysis of homologous and heterologous coverage curves indicated that the planktonic
274 and biofilm communities were significantly different from each other ($p = 0.001$). The
275 heterologous coverage of the planktonic library by the biofilm library was significantly
276 different ($p = 0.001$) from the homologous coverage of the planktonic library by itself
277 thus indicating that some sequences from the planktonic library had no close relatives in
278 the biofilm library (data not shown). This is supported by visual inspection of the
279 phylogenetic tree (Fig. 1). Homologous coverage, a measure of similarity within a given
280 library, was 100% for the biofilm library and 96% for the planktonic library.
281 Heterologous coverage, a measure of the representation of a given library in another, was
282 43% for the biofilm compared to the planktonic and 17% for the planktonic compared to
283 the biofilm at an evolutionary distance of 0.01 (99% 18S rDNA similarity OTU
284 definition).
285

286 **Species richness.** Non-parametric estimation of species richness calculated by two
287 different methods (ACE and Chao1) in DOTUR (67) revealed that the planktonic
288 community was more species-rich than the biofilm community (Table 1).
289
290 **Phylogenetic diversity and comparison of diversity indices.** Genetic differentiation
291 among the sampled microbial communities was assessed using the F_{ST} test (53). If there
292 is no effect of sample origin on phylogeny, the F_{ST} value should be close to zero (53).
293 The more pronounced the effect of sample origin on phylogeny, the more the value of F_{ST}
294 will deviate from zero. An F_{ST} value of 0.08497 ($p<0.001$) was obtained for the two
295 communities demonstrating that the levels of genetic diversity were significant, implying
296 that the within-community variation was different to the total diversity sampled when the
297 two communities were combined. This is in agreement with the ∫-LIBSHUFF results.
298 Since the choanoflagellates accounted for a significant portion of the biofilm population
299 and were also found in the planktonic community and since it appeared that these
300 sequences comprised distinct clades, it was of interest to determine how genetically
301 related the planktonic and biofilm choanoflagellate sequences were to each other. An F_{ST}
302 value of 0.50769 ($p<0.001$) was obtained for the comparison of the choanoflagellate
303 sequences between the two communities indicating that these populations were
304 genetically discrete between the planktonic and biofilm communities.
305 Various diversity indices were calculated for the planktonic and biofilm
306 eukaryotic libraries (Table 1). The reciprocal of Simpson's index is sensitive to the level
307 of dominance in a community and indicated that the planktonic eukaryotic community
308 (1/D index of 25) was five times more diverse than the biofilm community (1/D index of

309 5). The values for the Shannon-Weiner index, gene diversity and total genetic variation
310 show similar patterns (Table 1) with the planktonic community being more diverse than
311 the biofilm community which is also reflected in the phylogenetic distribution. The
312 planktonic library indeed contained representatives of more phyla than did the biofilm
313 library (Fig. 1).
314
315 **DISCUSSION**
316 We have compared the eukaryotic diversity of both the planktonic and biofilm
317 communities of the Snake River Plain Aquifer. Compared to prokaryotic diversity,
318 comparatively little is known about the eukaryotes which contribute to deep subsurface,
319 oxic aquifer habitats. While planktonic and sediment-associated protists have been
320 previously described in aquifers (38-40, 48), to our knowledge, this is the first report of a
321 molecular-based community assessment that includes biofilm-associated eukarya
322 sampled from an aquifer and the first to apply species richness and phylogenetic diversity
323 estimates of eukarya to an aquifer environment.
324 Rarefaction analysis was performed to determine the number of unique bacterial
325 clones as a proportion of the estimated total diversity. The rarefaction curve for the
326 planktonic-derived eukaryotic population did not reach a clear saturation indicating that
327 further sampling of this clone library may have revealed additional diversity. In contrast,
328 the rarefaction curve for the biofilm eukaryotic community reached a clear saturation
329 suggesting this library had been adequately sampled and that further sampling would not
330 likely reveal additional diversity. This is in agreement with non-parametric estimates for
331 both communities (Table 1).

332 Comparison of all diversity measures (Table 1) combined with the F_{ST} analysis
333 (0.08497, $p<0.0001$) showed that eukaryotic diversity in the planktonic fraction of the
334 SRPA is much higher than the eukaryotic biofilm diversity. The two eukaryotic
335 communities were also significantly different as determined by ∫-LIBSHUFF. These
336 observations are consistent with the findings from previous studies on attached (biofilm)
337 and planktonic prokaryotic communities in fractured rock aquifers. Lehman et al. (46)
338 found that planktonic and attached communities in an acidic crystalline rock aquifer were
339 distinct with respect to the biomass and the physiological capacities of the two fractions.
340 In the SRPA, prokaryotic communities present in water pumped from well TAN-33
341 (located approximately three kilometers from wells ANP-9 and ANP-10) were different
342 from prokaryotic communities obtained from basalt substrates that were incubated in the
343 well for eight months (45). Differences between planktonic communities in the attached
344 (biofilm) and planktonic phases of an aquifer may be related to the differences that we
345 observed in the eukarya present in these distinct phases.
346 The difference in diversity between the attached and planktonic communities may
347 also have been due to differences in how the two communities were sampled. For
348 planktonic samples, we concentrated the cells present in 13,300 L of groundwater. Given
349 that the basalts in the aquifer have a nominal porosity of 12% (17) this 13,300 L of
350 groundwater would have been obtained from a subsurface volume of approximately
351 40,000 L (or 40 m^3). By comparison, the basalt substrates that were incubated for six
352 months in ANP-10 in order to collect the attached biomass assayed a smaller volume of
353 the aquifer.

354 Novel clades belonging to the fungi were observed for both the biofilm and
355 planktonic communities: clade FI (EUKDBF10) representing 27% (10/37) of the
356 recovered biofilm sequences; clade FII (EPK B H1 and EPK B G4) and clade FIII (clones
357 EPK B G7, EUK DPK 65, EUKDPK41, and EPK B B11) representing 7% (6/86) and
358 19% (16/86) of the recovered planktonic sequences respectively. Novel
359 choanoflagellates (EBF C E3 and EBF C D10; EUKDPK38, EPK B E10, EPK B C12,
360 EPK B C11 and EPK B C4) were also observed in both communities. Novel clades
361 belonging to the stramenopiles were observed only in the planktonic population: clade SI
362 (EPK B G5 and EPK B F7) representing 8% (7/86), clade SII (EUKDPK45) representing
363 3% (3/86), clade III (EUKDPK52, EUKDPK44 and EUKDPK69) representing 10%
364 (9/86), clade SIV (EPK B A10 and EPK B C3) representing 2% (2/86), and clade SV
365 (EPK B C10) representing 7% (6/86) of the recovered planktonic sequences. It is of note
366 that clade SIII affiliated with an uncultured eukaryote clone isolated from Spain's River
367 of Fire (uncultured eukaryote clone Rt5ii25; AY082983). This clone was suggested to be
368 a member of a novel stramenopile lineage (2) and our results confirm this. Overall, 30%
369 of the planktonic community was composed of novel stramenopile sequences. Massana
370 et al. (54) observed that novel stramenopiles were found in all major marine habitats and
371 are important members of eukaryotic picoplankton. Our results suggest this is also true
372 for oligotrophic aquifer environments such as the SRPA. We did not observe any
373 stramenopile representatives within the biofilm population (54). These novel
374 environmental sequences significantly expand the known extent of eukaryotic rDNA
375 diversity within the deep subsurface. It is possible that some of these sequences may
376 represent eukaryotes identified in previous culture or microscopic studies but are not

represented by rRNA sequences in the database. Due to the paucity of available 18S rRNA sequence information, concrete conclusions with respect to the significance of these novel clades must be made with caution. It is interesting that 44 out of 81 eukaryotic biofilm-derived clones analyzed when sequenced, were determined to be archaeal in origin. Although archaea are present in the aquifer (63), this phenomenon was not observed with the planktonic-derived clones nor was it observed when two separate archaeal genomic templates were used as controls. The NS3/NS8 primer pair, allowing for a single mismatch will amplify archaeal representatives.

The planktonic and biofilm eukaryal communities were found to be significantly different (Fig. 2, Tables 1 and 2). The fungi dominated the planktonic eukaryal community (49%) while the choanoflagellates were the most abundant group within the biofilm community (70%). Cercozoa (16%, Fig. 2; Table 2) were only found in the biofilm population and were represented by a single clone (EUKDBF47) which was most closely affiliated with an uncultured cercozoan clone (AY620270) (6).

Choanoflagellates are generally considered to be significant members of marine and freshwater heterotrophic flagellate assemblages and are among the closest living single-celled relatives of metazoans (36) and may represent a model taxon for investigating metazoan origins (36, 37). That we found novel representatives within the choanoflagellates may contribute significantly to the investigation of metazoan origins. It is interesting that the planktonic choanoflagellate population are genetically distinct lineages from the biofilm choanoflagellates ($F_{ST} = 0.50769$, $p<0.001$). Whether or not this difference is due to selective feeding remains to be investigated. Heterotrophic flagellates such as the choanoflagellates have been recognized as consumers of

400 planktonic and attached (biofilm) bacteria (28, 44). Choanoflagellates are known to be
401 effective filter feeders and are thought to have an advantage over other flagellates when
402 food concentrations are low (7). In the deep basalt aquifer of the SRPA,
403 choanoflagellates dominate the recovered sequences from the biofilm (70%, Fig. 2) and a
404 significant portion of the recovered sequences from the planktonic community (19%, Fig.
405 2; Table 2).

406 A single clone (EUKDBF47) affiliated with the cercozoa made up 16% of the
407 biofilm-derived clones. Cercozoa are small flagellates present in many different
408 environments. It is likely that the organism represented by this clone (EUKDBF47) also
409 grazes on the biofilm-associated bacterial population.

410 The most common protists as identified by cultural and microscopic studies
411 reported for aquifer environments are flagellates and amoebae (62). While we did
412 observe flagellates, we did not observe amoebae. It is possible that the amoebae present
413 in the SRPA are in too low a concentration to be detected. ARB (49) analysis indicates
414 representatives of the Acanthamoebidae can be amplified by the NS3/NS8 primer pair.

415 A single fungal clone (EBF CD5) obtained from the biofilm was found to affiliate
416 with the genus *Emericella.* While a more diverse fungal population was found in the
417 planktonic sample, a planktonic clone (EUKDPK54) was also found to affiliate with
418 *Emericella.* An alignment of the two putative *Emericella* clones showed that they were
419 similar (96%) but not identical.

420 In the planktonic-derived eukaryote community, 13% of the clones analyzed
421 (11/86) were closely affiliated with *Exophiala* sp., dimorphic, black yeasts which are
422 members of the order Chaetothyriales and are causative agents of human mycoses (21).

A few noteable members of the genus *Exophiala*, such as *Exophiala lecanii-cornii* (15, 65, 77), *Exophiala jeanselmei* (65) and *Exophiala oligosperma* (27) have important bioremediation potential due to their ability to degrade a wide range of volatile organic compounds. It is noteworthy that these same species are also pathogenic (8, 20). The potential pathogenicity and bioremediation potential of the organisms within the SRPA represented by clones EUKDPK66 and EUKDPK64 cannot be delineated from their 18S signatures and thus future work to address this issue would be required.

It is not known whether the natural attenuation of the TCE contaminated plume of the SRPA occurs within the planktonic population, the biofilm population or both. Knowledge of the eukaryal populations that coexist with prokaryotic populations is vital to the future of the study of natural attenuation in the deep subsurface of fractured basalt aquifers such as the SRPA. The eukaryal populations could potentially impact remediation efforts in a positive manner. For instance, *Exophiala* (27, 65, 77) and other fungi with the capacity for degradation in addition to protists which are able to uptake contaminants (40, 42, 71) may enhance intrinsic bioremediation rates as will predation by protists which has been demonstrated to increase uptake by degradative bacteria (34, 70). By the same token, the eukaryal populations may negatively impact remediation efforts by over-grazing of contaminant degrading bacteria (and potentially Archaea). The delicate balance of indigenous microbes within contaminated and uncontaminated subsurface environments must be understood. Knowledge gained by investigations such as the one presented here will facilitate interpretation of diversity studies within contaminated aquifer environments.

AKNOWLEDGEMENTS

This research was supported by the U.S Department of Energy, Grant No. DE-FG07-02ER63500. Research at the Idaho National Laboratory now operated by Battelle Energy Alliance, under Contract DE-AC07-05ID14517, was supported by the U.S. Department of Energy, Office of Science, Environmental Management Science Program. Phylogenetic analysis on the cluster computer was made possible by grants from NSF EPSCOR EPS 0080935, NIH INBRE P20 RR16454 and NIH COBRE P20 RR16448. We thank Mark Delwiche for sampling efforts, Daniel Erwin for sampling and total DNA extraction, Celeste Brown for assistance with the cluster computer and Cornelia Sawatzky for editorial assistance.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403-10.

2. **Amaral Zettler, L. A., G. F., E. Zettler, B. G. Keenan, R. Amils, and M. L. Sogin.** 2002. Eukaryotic diversity in Spain's River of Fire. Nature **417:**137.

3. **Amaral Zettler, L. A., M. A. Messerli, A. D. Laatsch, P. J. S. Smith, and M. L. Sogin.** 2003. From genes to genomes: beyond biodiversity in Spain's Rio Tinto. Biol. Bull. **204:**205-209.

4. **Arndt, H., D. Dietrich, B. Auer, E.-J. Cleven, T. Grafenhan, M. Weitere, and A. P. Mylnikov.** 2000. Functional diversity of heterotrophic flagellates in aquatic ecosystems. *In* B. S. C. Leadbeater and J. C. Green (ed.), The flagellates: unity, diversity and evolution. Taylor and Francis, London.

5. **Baker, B. J., M. A. Lutz, S. C. Dawson, P. L. Bond, and J. F. Banfield.** 2004. Metabolically active eukaryotic communities in extremely acidic mine drainage. Appl. Environ. Microbiol. **70:**6264-6271.

6. **Bass, D., and T. Cavalier-Smith.** 2004. Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). Int. J. Syst. Evol. Microbiol. **54:**2393-404.

7. **Boenigk, J., and H. Arndt.** 2002. Bacterivory by heterotrophic flagellates: community structure and feeding strategies. Antonie Van Leeuwenhoek **81:**465-80.

8. **Bossler, A. D., S. S. Richter, A. J. Chavez, S. A. Vogelgesang, D. A. Sutton, A. M. Grooters, M. G. Rinaldi, G. S. de Hoog, and M. A. Pfaller.** 2003.

Exophiala oligosperma causing olecranon bursitis. J. Clin. Microbiol. **41:**4779-4782.

9. **Buchan, A., S. Y. Newell, M. Butler, E. J. Biers, J. T. Hollibaugh, and M. A. Moran.** 2003. Dynamics of bacterial and fungal communities on decaying salt marsh grass. Appl. Environ. Microbiol. **69:**6676-6687.

10. **Caron, D. A., R. J. Gast, E. L. Lim, and M. R. Dennett.** 1999. Protistan community structure: molecular approaches for answering ecological questions. Hydrobiologia **401:**215-227.

11. **Casamayor, E. O., R. Massana, S. Benlloch, L. Ovreas, B. Diez, V. J. Goddard, J. M. Gasol, I. Joint, F. Rodriguez-Valera, and C. Pedros-Alio.** 2002. Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. Environ. Microbiol. **4:**338-48.

12. **Castresana, J.** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17:**540-552.

13. **Chao, A. 1987.** Estimating the population-size for capture recapture data with unequal catchability. Biometrics **43:**783-791.

14. **Chao, A., and S. M. Lee.** 1992. Estimating the number of classes via sample coverage. J. Am. Stat. Assoc. **87:**210-217.

15. **Cheng, Y. Q., G. L. Tang, and B. Shen.** 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. Proc. Natl. Acad. Sci. U.S.A. **100:**3149-54.

16. **Christner, B. C., B. H. Kvitko, and J. N. Reeve. 2003.** Molecular identification of bacteria and eukarya inhabiting an Antarctic cryoconite hole. Extremophiles **7:**1425-1430.

17. **Colwell, F. S., R. W. Smith, F. G. Ferris, A.-L. Reysenbach, Y. Fujita, T. L. Tyler, J. L. Taylor, A. Banta, M. E. Delwiche, T. McLing, and M. E. Watwood.** 2005. Microbially-mediated subsurface calcite precipitation for removal of hazardous divalent cations: Microbial activity, molecular biology, and modeling, p. 117-137. *In* E. Berkey and T. Zachary (ed.), Subsurface Contamination Remediation: Accomplishments of the Environmental Management Science Program, vol. 904. American Chemical Society, Washington, DC.

18. **Countway, P. D., R. J. Gast, P. Savai, and D. A. Caron.** 2005. Protistan diversity estimates based on 18S rDNA from seawater incubations in the Western North Atlantic. J. Eukaryot. Microbiol. **52:**95-106.

19. **Dawson, S. C., and N. R. Pace.** 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. Proc. Natl. Acad. Sci. U.S.A. **99:**8324-8329.

20. **De Hoog, G. S., B. Bowman, Y. Graser, G. Haase, M. El Fari, A. H. Gerrits van den Ende, B. Melzer-Krick, and W. A. Untereiner.** 1998. Molecular phylogeny and taxonomy of medically important fungi. Med. Mycol. **S1:**52-6.

21. **de Hoog, G. S., V. Vicente, R. B. Caligiorne, S. Kantarcioglu, K. Tintelnot, A. H. G. Gerrits van den Ende, and G. Haase.** 2003. Species diversity and polymorphism in the *Exophiala spinifera* clade containing opportunistic black yeast-like fungi. J. Clin. Microbiol. **41**:4767-4778.

22. **de Souza, F. A., G. A. Kowalchuk, P. Leeflang, J. A. van Veen, and E. Smit.** 2004. PCR-denaturing gradient gel electrophoresis profiling of inter-and intraspecies 18S rRNA gene sequence heterogeneity is an accurate and sensitive method to assess species diversity of arbuscular mycorrhizal fungi of the genus *Gigaspora*. Appl. Environ. Microbiol. **70:**1413-1424.

23. **Diez, B., C. Pedros-Alio, and R. Massana.** 2001. Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. Appl. Environ. Microbiol. **67:**2932-2941.

24. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. **32:**1792-1797.

25. **Ekendahl, S., A. H. O'Neill, E. Thomsson, and K. Pedersen.** 2003. Characterisation of yeasts isolated from deep igneous rock aquifers of the Fennoscandian Shield. Microb. Ecol. **46:**416-28.

26. **Erwin, D. P., I. K. Erickson, M. E. Delwiche, F. S. Colwell, J. L. Strap, and R. L. Crawford.** 2005. Diversity of oxygenase genes from methane- and ammonia-oxidizing bacteria in the Eastern Snake River Plain Aquifer. Appl. Environ. Microbiol. **71:**2016-2025.

27. **Estevez, E., M. C. Veiga, and C. Kennes.** 2005. Biodegradation of toluene by the new fungal isolates *Paecilomyces variotii* and *Exophiala oligosperma*. J. Ind. Microbiol. Biotechnol. **32:**33-7.

28. **Fenchel, T.** 1986. The ecology of heterotrophic flagellates, p. 55-97. *In* K. C. Marshall (ed.), Advances in Microbial Ecology. Plenum Press, New York.

29. **Galand, P. E., S. Saarnio, H. Fritze, and K. Yrjala.** 2002. Depth related diversity of methanogen Archaea in Finnish oligotrophic fen. FEMS Microbiol. Ecol. **42:**441-449.

30. **Good, I. J.** 1953. The population frequency of species and the estimation of population parameters. Biometrika **40:**237-264.

31. **Hackl, E., S. Zechmeister-Boltenstern, L. Bodrossy, and A. Sessitsch.** 2004. Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils. Appl. Environ. Microbiol. **70:**5057-5065.

32. **Huber, T., G. Faulkner, and P. Hugenholtz.** 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics **20:**2317-9.

33. **Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. M. Bohannan.** 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. Appl. Environ. Microbiol. **67:**4399-4406.

34. **Hunt, H. W., C. V. Cole, D. A. Klein, and D. C. Coleman.** 1977. A simulation for the effect of predation on bacteria in continuous culture. Microb. Ecol. **3:**259-278.

35. **Huws, S. A., A. J. McBain, and P. Gilbert.** 2005. Protozoan grazing and its impact upon population dynamics in biofilm communities. J. Appl. Microbiol. **98:**238-244.

36. **King, N.** 2005. Choanoflagellates. Curr. Biol. **15:**R113-R114.

37. **King, N.** 2004. The unicellular ancestry of animal development. Dev. Cell **7:**313-25.

38. **Kinner, N. E., R. W. Harvey, K. Blakeslee, G. Novarino, and L. D. Meeker.** 1998. Size-selective predation on groundwater bacteria by nanoflagellates in an organic-contaminated aquifer. Appl. Environ. Microbiol. **64:**618-625.

39. **Kinner, N. E., R. W. Harvey, and M. Kazmierkiewicz-Tabaka.** 1997. Effect of flagellates on free-living bacterial abundance in an organically contaminated aquifer. FEMS Microbiol. Rev. **20:**249-59.

40. **Kinner, N. E., R. W. Harvey, D. M. Shay, D. W. Metge, and A. Warren.** 2002. Field evidence for a protistan role in an organically-contaminated aquifer. Environ. Sci. Technol. **36:**4312-8.

41. **Kuhlman, K. R., W. G. Fusco, M. T. La Duc, L. B. Allenbach, C. L. Ball, G. M. Kuhlman, R. C. Anderson, I. K. Erickson, T. Stuecker, J. Benardini, J. L. Strap, and R. L. Crawford.** 2005. Diversity of microorganisms within rock varnish in the Whipple Mountains, California. Appl. Environ. Microbiol. (In Press).

42. **Kujawinski, E. B., J. W. Farrington, and J. W. Moffett.** 2000. Importance of passive diffusion in the uptake of polychlorinated biphenyls by phagotrophic protozoa. Appl. Environ. Microbiol. **66:**1987-1993.

43. **Lawley, B., S. Ripley, P. Bridge, and P. Convey.** 2004. Molecular Analysis of Geographic Patterns of Eukaryotic Diversity in Antarctic Soils. Appl. Environ. Microbiol. **70:**5963-5972.

44. **Laybourn-Parry, J., and J. Parry.** 2000. Flagellates and the microbial loop, p. 216-239. *In* B. S. C. Leadbeater and J. C. Green (ed.), The Flagellates. Taylor and Francis, London.

45. Lehman, R. M., S. P. O'Connell, A. Banta, J. Fredrickson, A.-L. Reysenbach, T. L. Kieft, and F. S. Colwell. 2004. Microbiological comparison of core and groundwater samples collected from a fractured basalt aquifer with that of dialysis chambers incubated *in situ*. Geomicrobiol. J. **21:**169-182.

46. Lehman, R. M., F. F. Roberto, D. Earley, D. F. Bruhn, S. E. Brink, S. P. O'Connell, M. E. Delwiche, and F. S. Colwell. 2001. Attached and unattached bacterial communities in a 120-meter corehole in an acidic, crystalline rock aquifer. Appl. Environ. Microbiol. **67:**2095-2106.

47. López-Garcia, P., F. Rodriguez-Valera, C. Pedros-Alio, and D. Moreira. 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature **409:**603-7.

48. Ludvigsen, L., H. J. Albrechtsen, D. B. Ringelberg, F. Ekelund, and T. H. Christensen. 1999. Distribution and composition of microbial populations in a landfill leachate contaminated aquifer (Grindsted, Denmark). Microb. Ecol. **37:**197-207.

49. Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Y. Kumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüßmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.-H. Schleifer. 2004. ARB: a software environment for sequence data. Nucl. Acids Res. **32:**1363-1371.

50. **Madsen, E. L., J. L. Sinclair, and W. C. Ghiorse.** 1991. In situ biodegradation: microbiological patterns in a contaminated aquifer. Science **252:**830-3.

51. **Maidak, B. L., J. R. Cole, T. G. Lilburn, C. T. Parker, Jr., P. R. Saxman, R. J. Farris, G. M. Garrity, G. J. Olsen, T. M. Schmidt, and J. M. Tiedje.** 2001. The RDP-II (Ribosomal Database Project). Nucl. Acids Res. **29:**173-4.

52. **Maidak, B. L., G. J. Olsen, N. Larsen, R. Overbeek, M. J. McCaughey, and C. R. Woese.** 1996. The Ribosomal Database Project (RDP). Nucl. Acids Res. **24:**82-5.

53. **Martin, A. P.** 2002. Phylogenetic approaches for describing and comparing the diversity of microbial communities. Appl. Environ. Microbiol. **68:**3673-3682.

54. **Massana, R., J. Castresana, V. Balague, L. Guillou, K. Romari, A. Groisillier, K. Valentin, and C. Pedros-Alio.** 2004. Phylogenetic and ecological analysis of novel marine stramenopiles. Appl. Environ. Microbiol. **70:**3528-34.

55. **Massana, R., L. Guillou, B. Diez, and C. Pedros-Alio.** 2002. Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. Appl. Environ. Microbiol. **68:**4554-4558.

56. **Matz, C., and K. Jurgens.** 2003. Interaction of nutrient limitation and protozoan grazing determines the phenotypic structure of a bacterial community. Microb. Ecol. **45:**384-398.

57. **Minin, V., Z. Abdo, P. Joyce, and J. Sullivan.** 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. **52:**674-683.

636 **58.** **Moon-van der Staay, S. Y., R. De Wachter, and D. Vaulot.** 2001. Oceanic 18S
637 rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity.
638 Nature **409:**607-10.
639 **59.** **Mullins, T. D., T. B. Britschgi, R. I. Krest, and S. J. Giovannoni.** 1995.
640 Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and
641 Pacific bacterioplankton communities. Limnol. Oceanogr. **39:**148-158.
642 **60.** **Newby, D. T., D. W. Reed, L. M. Petzke, A. L. Igoe, M. E. Delwiche, F. F.**
643 **Roberto, J. P. McKinley, M. J. Whiticar, and F. S. Colwell.** 2004. Diversity of
644 methanotroph communities in a basalt aquifer. FEMS Microbiol. Lett. **48:**333-
645 344.
646 **61.** **Nikolcheva, L., A. Cockshutt, and F. Barlocher.** 2003. Determining diversity of
647 freshwater fungi on decaying leaves: comparison of traditional and molecular
648 approaches. Appl. Environ. Microbiol. **69:**2548-2554.
649 **62.** **Novarino, G., A. Warren, H. Butler, G. Lambourne, A. Boxshall, J. Bateman,**
650 **N. E. Kinner, R. W. Harvey, R. A. Mosse, and B. Teltsch.** 1997. Protistan
651 communities in aquifers: a review. FEMS Microbiol. Rev. **20:**261-75.
652 **63.** **O'Connell, S. P., R. M. Lehman, O. Snoeyenbos-West, V. D. Winston, D. E.**
653 **Cummings, M. E. Watwood, and F. S. Colwell.** 2003. Detection of
654 *Euryarchaeota* and *Crenarchaeota* in an oxic basalt aquifer. FEMS Microbiol.
655 Ecol. **44:**165-173.
656 **64.** **Page, R.** 1996. TreeView: an application to display phylogenetic trees on
657 personal computers. Computer Appl. Biosci. **12:**357-358.

65. **Qi, B., W. M. Moe, and K. A. Kinney.** 2002. Biodegradation of volatile organic compounds by five fungal species. Appl. Microbiol. Biotechnol. **58:**684-9.

66. **Savin, M. C., J. L. Martin, M. LeGresley, M. Giewat, and J. Rooney-Varga.** 2004. Plankton diversity in the Bay of Fundy as measured by morphological and molecular methods. Microb. Ecol. **48:**51-65.

67. **Schloss, P. D., and J. Handelsman.** 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. **71:**1501-1506.

68. **Schloss, P. D., B. R. Larget, and J. Handelsman.** 2004. Integration of microbial ecology and statistics: a test to compare gene libraries. Appl. Environ. Microbiol. **70:**5485-92.

69. **Schneider, S., D. Roessli, and L. Excoffier.** 2000. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, Univerisity of Geneva, Switzerland.

70. **Sherr, B. F., E. B. Sherr, and T. Berman.** 1983. Grazing, growth, and ammonium excretion rates of a heterotrophic microflagellate fed with four speices of bacteria. Appl. Environ. Microbiol. **45:**1196-1201.

71. **Sherr, E. B.** 1988. Direct use of high molecular weight polysaccharide by heterotrophic flagellates. Nature **335:**348-351.

72. **Sorenson, K. S., Jr., L. N. Peterson, R. E. Hinchee, and R. Ely.** 2000. An evaluation of aerobic trichloroethene attenuation using first-order rate estimation. Bioremed. J. **4:**337-357.

73. **Stoeck, T., G. T. Taylor, and S. S. Epstein.** 2003. Novel eukaryotes from the permanently anoxic Cariaco Basin (Caribbean Sea). Appl. Environ. Microbiol. **69:**5656-5663.

74. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. **25:**4876-82.

75. **Weisse, T.** 2002. The significance of inter- and intraspecific variation in bacterivorous and herbivorous protists. Antonie Van Leeuwenhoek **81:**327-341.

76. **White, T. J., T. Bruns, S. Lee, and J. Taylor.** 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p. 315-322. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR Protocols: A Guide to Methods and Applications. Academic Press, Inc., New York.

77. **Woertz, J. R., K. A. Kinney, N. D. McIntosh, and P. J. Szaniszlo.** 2001. Removal of toluene in a vapor-phase bioreactor containing a strain of the dimorphic black yeast *Exophiala lecanii-corni*. Biotechnol. Bioeng. **75:**550-8.

78. **Zhou, J., B. Xia, D. S. Treves, L.-Y. Wu, T. L. Marsh, R. V. O'Neill, A. V. Palumbo, and J. M. Tiedje.** 2002. Spatial and resource factors influencing high microbial diversity in soil. Appl. Environ. Microbiol. **68:**326-334.

79. **Zubkov, M. V., and M. A. Sleigh.** 2000. Comparison of growth efficiencies of protozoa growing on bacteria deposited on surfaces and in suspension. J. Eukaryot. Microbiol. **47:**62-69.

703 Table 1. Comparison of diversity indices for the planktonic and biofilm communities

704 sampled from the SRPA. (Values in parentheses are the 95% confidence intervals)

705
706

Community	Planktonic	Biofilm
No. Non-Chimeric Clones	86	37[a]
No. Distinct Sequences (RFLP)	24	5
No. OTUs (≥99%)	24	5
ACE [b]	26	5
Chao1	30 (25, 72)	5
Shannon-Weiner Index	3.05 (2.92, 3.17)	1.58 (1.46, 1.69)
1/D[c]	25	5
Nucleotide Diversity	0.432 ± 0.213	0.229 ± 0.139
Gene Diversity	0.993 ±0.014	0.900 ± 0.161
θ(π) [d]	659.2 ± 325.1	329.4 ± 199.9

707

708 [a] 81 clones were initially analyzed, of these 44 were found to be archaea

709 [b] Abundance-based richness estimator

710 [c] Reciprocal of Simpson's index (D)

711 [d] Total Genetic Variation

712
713

Table 2. BLAST Analysis of 18S rDNA sequences amplified from planktonic (PK, n = 86) and biofilm (BF, n = 37) samples from the SRPA

Clone ID	Accession No.	Frequency	Taxonomic Affiliation	Nearest Phylogenetic Neighbor	Accession No.	% Similarity[a]	Identities
EBF C D5	DQ104580	5	Fungi/Ascomycota/ Eurotiomycetes/ Eurotiales	*Emericella nidulans*	AB008403	99	1182/1191
EBF C D10	DQ104581	9	Fungi/Metazoa group/Choanoflagellida	*Monosiga brevicollis*	AF174375	92	968/1049
EBF C E3	DQ104582	7	Fungi/Metazoa group/Choanoflagellida	*Monosiga ovata*	AF084230	93	887/944
EUKDBF10	DQ104583	10	Fungi/Chytridiomycota/Chytridiales	*Lacustromyces hiemalis*	AF164274	93	184/196
EUKDBF47	DQ104584	6	Cercozoa	Uncultured cercozoan clone 4-6.2	AY620270.1	97	709/728
EUKDPK69	DQ104585	5	unclassified	Uncultured eukaryote clone RT5iin25	AY082983.1	95	253/264
EPK B B11	DQ104586	5	Fungi/Ascomycota/Saccharomycetes/Saccharomycetales	*Candida tanzawaensis* strain NRRL Y-17324	AY227713.1	94	265/281
EPK B C4	DQ104587	3	Fungi/Metazoa group/Choanoflagellida	*Monosiga ovata*	AF084230.1	93	884/943
EPK B C10	DQ104588	6	stramenopiles/Bicosoecida/Caecitellus	*Caecitellus parvulus* strain NBH4	AF174368	91	556/605
EPK B C11	DQ104589	1	Fungi/Metazoa group/Choanoflagellida	*Monosiga brevicollis*	AF174375	92	965/1046
EPK B C12	DQ104590	4	Fungi/Metazoa group/Choanoflagellida	*Monosiga ovata*	AF084230.1	93	568/608
EPK B E10	DQ104591	4	Fungi/Metazoa group/Choanoflagellida	*Monosiga ovata*	AF084230.1	96	576/594
EPK B F7	DQ104592	5	no match				
EPK B G4	DQ104593	4	no match				
EPK B G5	DQ104594	2	no match				
EPK B G7	DQ104595	3	unclassified	Uncultured fungus clone LEM108	AY168908	93	763/817
EPK B H1	DQ104596	2	no match				
EUKDPK38	DQ104597	4	Fungi/Metazoa group/Choanoflagellida	*Monosiga ovata*	AF271999.1	93	712/765
EUKDPK40	DQ104598	5	Fungi/Metazoa group/Ascomycota/Dothideomycetes/Pleosporales	*Paraphaeosphaeria* sp. N119	AB096264.1	91	1134/1235
EUKDPK41	DQ104599	3	unclassified	Uncultured fungus clone LEM108	AY168908	93	1001/1074
EUKDPK44	DQ104600	1	unclassified	Uncultured eukaryote clone RT5iin25	AY082983.1	96	254/262
EUKDPK45	DQ104601	3	unclassified	Uncultured fungus clone 21_14	AY168908.1	93	389/414
EUKDPK52	DQ104602	3	unclassified	Uncultured eukaryote clone RT5iin25	AY082983.1	96	253/262
EUKDPK54	DQ104603	7	Fungi/Ascomycota/Eurotiomycetes/Eurotiales	*Emericella nidulans*	AB008403.1	97	1202/1235
EUKDPK64	DQ104604	4	Fungi/Ascomycota/Chaetothyriomycetes/Chaetothyriales	*Exophiala* sp. strain GHP 1205	AJ232954.1	97	841/867
EUKDPK65	DQ104605	5	unclassified	Uncultured fungus clone LEM108	AY168908	95	541/565
EUKDPK66	DQ104606	5	Fungi/Ascomycota/Chaetothyriomycetes/Chaetothyriales	*Exophiala* sp. 122071	AF542179	96	419/435
EPK B C3	DQ104607	1	no match				
EPK B A10	DQ104608	1	no match				

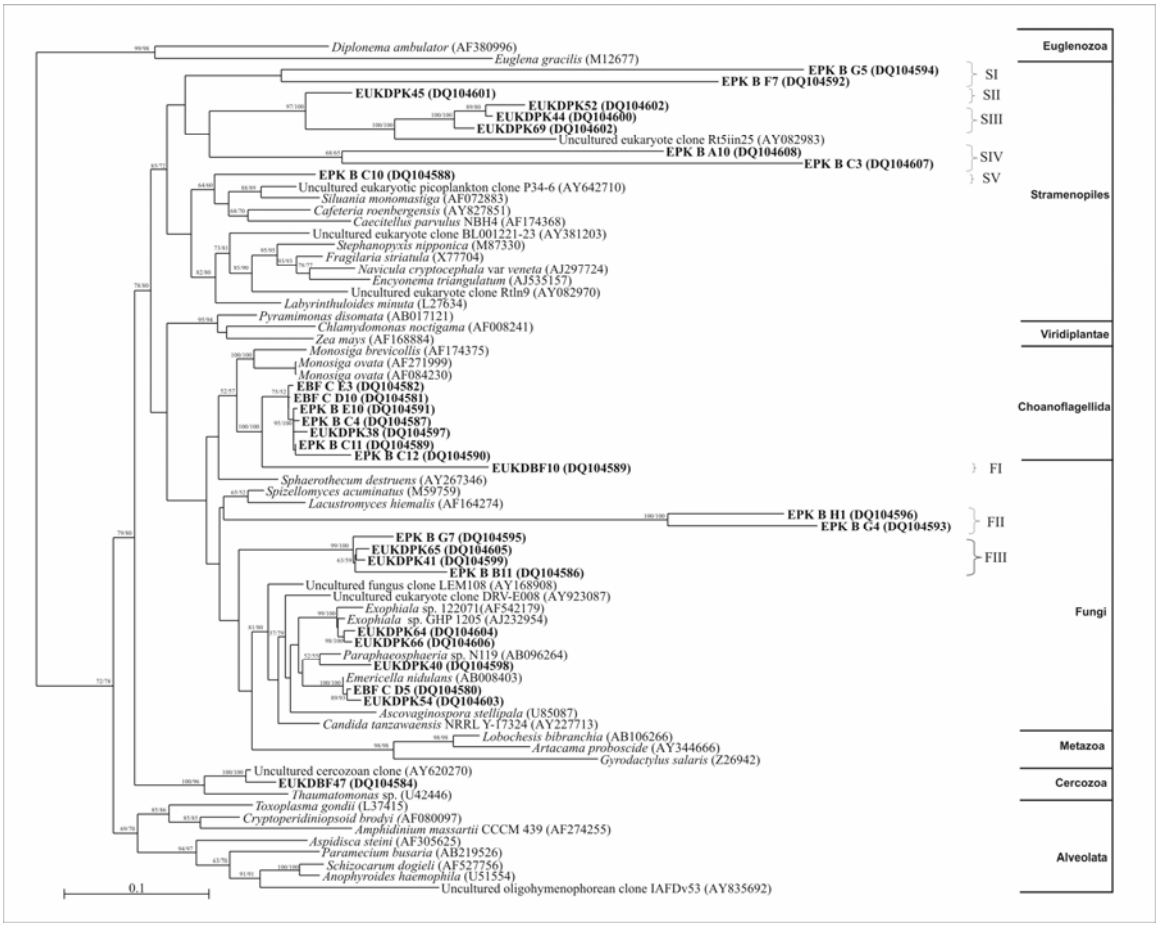
[a] A BLAST hit with identity <150 bp was considered 'no match'

Figure Legends

Fig. 1. Phylogenetic relationships among partial eukaryal 18S rDNA sequences amplified from the planktonic (n = 86) and the biofilm (n = 37) community sampled within the SRPA. The bar represents 10% sequence divergence. The Maximum Likelihood (ML) tree was constructed using PAUP* from 1307 base positions using the TrNef+G correction model. Bootstrap values over 50% from an analysis of 1,000 bootstrap replicates are given at the respective nodes. Maximum Parsimony (MP) bootstrap values are also shown on the tree (ML/MP). Environmental clones obtained in this study are shown in bold. Clone designations containing BF were sampled from the biofilm while clone designations containing PK were from planktonic samples. GenBank accession numbers are listed in parentheses. Putatively novel clades are indicated by Roman numerals.

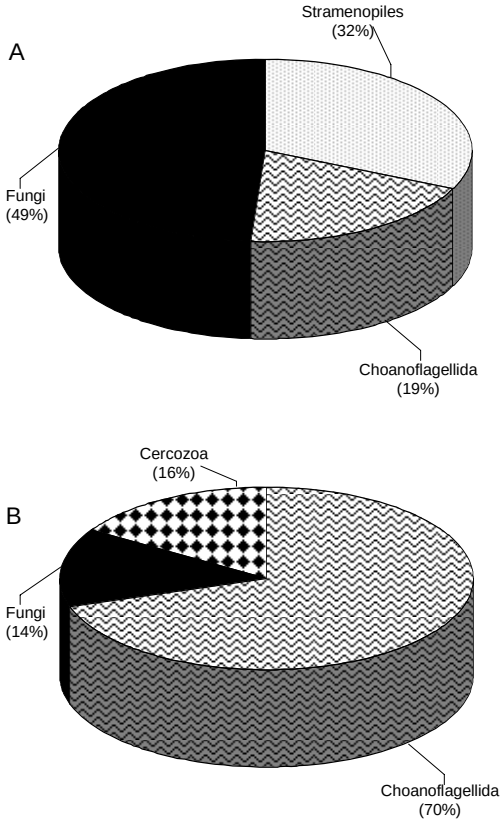
Fig. 2. Relative abundance of unique eukaryotic phylotypes within represented phylogenetic groups identified in planktonic (A) and biofilm (B) 18S rDNA libraries of the SRPA. Planktonic, n = 86; Biofilm, n = 37.

743 Figure 1



744

745



746 Figure 2
747

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Apr. 2005, p. 2016–2025 Vol. 71, No. 4
0099-2240/05/$08.00+0 doi:10.1128/AEM.71.4.2016–2025.2005

Diversity of Oxygenase Genes from Methane- and Ammonia-Oxidizing Bacteria in the Eastern Snake River Plain Aquifer

Daniel P. Erwin,[1] Issac K. Erickson,[2] Mark E. Delwiche,[3] Frederick S. Colwell,[3] Janice L. Strap,[1] and Ronald L. Crawford[1]*

Environmental Biotechnology Institute[1] and Biology Department,[2] University of Idaho, Moscow, and Idaho National Engineering and Environmental Laboratory, Idaho Falls,[3] Idaho

Received 30 July 2004/Accepted 10 November 2004

PCR amplification, restriction fragment length polymorphism, and phylogenetic analysis of oxygenase genes were used for the characterization of in situ methane- and ammonia-oxidizing bacteria from free-living and attached communities in the Eastern Snake River Plain aquifer. The following three methane monooxygenase (MMO) PCR primer sets were used: A189-A682, which amplifies an internal region of both the *pmoA* gene of the MMO particulate form and the *amoA* gene of ammonia monooxygenase; A189-mb661, which specifically targets the *pmoA* gene; and mmoXA-mmoXB, which amplifies the *mmoX* gene of the MMO soluble form (sMMO). Whole-genome amplification (WGA) was used to amplify metagenomic DNA from each community to assess its applicability for generating unbiased metagenomic template DNA. The majority of sequences in each archive were related to oxygenases of type II-like methanotrophs of the genus *Methylocystis*. A small subset of type I sequences found only in free-living communities possessed oxygenase genes that grouped nearest to *Methylobacter* and *Methylomonas* spp. Sequences similar to that of the *amoA* gene associated with ammonia-oxidizing bacteria (AOB) most closely matched a sequence from the uncultured bacterium BS870 but showed no substantial alignment to known cultured AOB. Based on these functional gene analyses, bacteria related to the type II methanotroph *Methylocystis* sp. were found to dominate both free-living and attached communities. Metagenomic DNA amplified by WGA showed characteristics similar to those of unamplified samples. Overall, numerous sMMO-like gene sequences that have been previously associated with high rates of trichloroethylene cometabolism were observed in both free-living and attached communities in this basaltic aquifer.

Trichloroethylene (TCE) is a common groundwater contaminant of toxic and carcinogenic concern (2, 32, 37, 48). Its extensive use as an industrial solvent, combined with inadequate disposal methods, has led to groundwater contamination of many sites worldwide. TCE itself is a suspected carcinogen, and its biotransformation product vinyl chloride, which can be produced in anaerobic aquifers, is a known carcinogen for animals (33, 35). Due to the established toxicity of this product, the Environmental Protection Agency has a nonenforceable maximum contaminant level goal of 0 μg liter^{-1} for TCE in drinking water supplies. Presently, the enforced maximum contaminant level for TCE is 5 μg liter^{-1} (16).

In 1989, Test Area North (TAN), located on the Idaho National Environmental and Engineering Laboratory site in southeastern Idaho, was added to the National Priorities List as one of the sites requiring cleanup (15). The TAN site contains a contamination plume within the Eastern Snake River Plain Aquifer (ESRPA) containing as much as 300 mg of TCE liter^{-1} and extending over an area of 25,000 m^2 (46). Wells have been monitored throughout the TAN site in order to reveal the boundaries of this TCE plume. Though the extent of this plume is quite large, it does not appear to be migrating farther from the initial point of contamination. A previous study ruled out abiotic loss as the only reason for this observed attenuation (47). In addition, direct evidence that TCE cometabolism is occurring at the TAN site within the plume of TCE contamination has been reported (M. H. Howard-Jones, W. K. Keener, R. A. Wymore, F. S. Colwell, K. S. Sorenson, and M. E. Watwood, Abstr. 103rd Gen. Meet. Am. Soc. Microbiol., abstr. Q-005, p. 514, 2003).

Previously, methanotrophs, propanotrophs, nitrifiers, and phenol oxidizers have been isolated from groundwater within the TAN site (7). Expression of a wide array of nonspecific oxygenase enzymes would allow these bacteria to use a variety of primary substrates that support cometabolic TCE dechlorination. Of these organisms, the methanotrophs in particular have been well studied for their ability to cometabolize groundwater pollutants (1, 9, 24, 31, 42, 49, 50).

Methanotrophs use methane monooxygenase (MMO) to catalyze the oxidation of their primary growth substrate methane to methanol. The enzyme can be found either in the cytoplasm in a soluble form (sMMO) or in the membrane-bound particulate form (pMMO) (17). The pMMO enzyme is found in all known methanotrophs, with the potential exception of an acidophilic methanotroph isolated by Dedysh et al. (13). Of the organisms also capable of sMMO production, the majority have been found to be of the type II and type X methanotrophs affiliated with the α subclass of *Proteobacteria* (17). Though rare, strains of type I methanotrophs (affiliated with γ-*Proteobacteria*) have also been found to produce sMMO (5, 39). The *mmoX* gene encodes the conserved α subunit of the hydroxylase component of the sMMO (6) and has been used previously as a marker for sMMO (36, 45).

Both the soluble and particulate forms of MMO have established potentials for the aerobic dechlorination of pollutants

* Corresponding author. Mailing address: University of Idaho, Environmental Biotechnology Institute, Food Research Center 204, Moscow, ID 83844-1052. Phone: (208) 885-6580. Fax: (208) 885-5741. E-mail: crawford@uidaho.edu.

such as TCE, though at different rates (3, 4). The sMMO enzyme is of particular interest to bioremediation studies, as it has been shown to exhibit a broader substrate range than the particulate form. In addition, the sMMO exhibits a higher level of activity than pMMO, which leads to a significantly higher rate of TCE degradation than that of pMMO (4, 10).

Ammonia monooxygenase (AMO) is similar to pMMO in both structure and function. AMO is used to obtain energy for carbon dioxide fixation through the oxidation of ammonia and is found in ammonia-oxidizing bacteria (AOB), such as *Nitrosomonas* and *Nitrosospira* spp., that group within the β subclass of *Proteobacteria* (17, 22). The genes encoding AMO and those encoding pMMO share high sequence identity and are believed to be evolutionarily related. This concept is supported through predicted amino acid sequence alignments showing that both primary and secondary structures are well conserved between subunits of both enzymes (19). For example, *Nitrosomonas europaea* can oxidize TCE through the action of AMO. The affinity of TCE for AMO is very close to its natural substrate, ammonia (22).

The objective of this study was to use molecular techniques to examine the presence and diversity of oxygenase genes from methane-oxidizing bacteria (MOB) and ammonia-oxidizing bacteria within a pristine region of the TAN site of the Idaho National Environmental and Engineering Laboratory, focusing on MMO and AMO genes of the total microbial community (free-living and attached). This information will assist us in understanding the responses of the natural microflora of the pristine aquifer to the appearance of contaminants such as TCE. Beyond simply confirming the presence of genes that enable TCE cometabolism, our aim was to assay the diversity of these genes as a first step in understanding how different versions of the same class of degradative enzymes might collectively contribute to cometabolism.

MATERIALS AND METHODS

Sampling. The microbial communities present in the ESRPA were sampled in the fall of 2003. Two wells, ANP-9 and ANP-10, located in a pristine region of the aquifer proximal to the TCE-contaminated plume, were used for this study. A detailed site description has recently been reported by Newby et al. (40). The close proximity of the wells to each other (ca. 200 m) suggests that their respective water chemistries and microbial community structures should be similar. Previous analysis indicates comparable water chemistry characteristics between the two wells and values that are typical of the ESRPA. For example, selected water chemistry values for ANP-9 and this region of the aquifer in general include temperature of 14.1 to 14.5°C, a pH of 7.2 to 7.9, a dissolved oxygen level of 7.5 to 8.4 mg liter^{-1}, a dissolved organic carbon level of 1.5 to 2.6 mg liter^{-1}, and a methane concentration of 72 nmol liter^{-1} (40). Well ANP-9 was used in the collection of free-living cells, while well ANP-10 was used for attached community sampling.

The free-living community in this study was defined as all cells freely dispersed within the aquifer groundwater or only loosely associated with the basalt. The attached community consisted of all cells that colonized the surfaces of basalt substrates that were incubated in well ANP-10. Methods were designed to facilitate the sampling of each community by using existing wells (18, 27-29).

A KrosFlo Pilot hollow-fiber, tangential flow filtration system with a DynaFibre microporous membrane (0.2 μm pore size and 3.9 m^2 in surface area; Spectrum Labs, Rancho Dominguez, Calif.) was employed for the collection of free-living organisms from ANP-9. Prior to the initiation of filtration, the well was flushed with 3 well volumes (1 volume equal to the standing water in the well) to remove built-up sediments and debris from the casing. Approximately 13,300 liters of aquifer water was then passed through the hollow-fiber apparatus and concentrated to approximately 3 liters. The concentrate was collected by reversing the pump motor and directing the retentate into sterile 1-liter containers. This concentrate was then stored at −80°C and thawed at room temperature prior to being used.

TABLE 1. Primer sequences used for the molecular characterization of microbial communities derived from the Eastern Snake River Plain Aquifer

Primer	Sequence (5′ to 3′)	Target	Reference
A189	GGNGACTGGGACTTCTGG	*amoA*/*pmoA*	18
A682	GAASGCNGAGAAGAASGC	*amoA*/*pmoA*	18
mb661	CCGGMGCAACGTCYTTACC	*pmoA*	10
mmoXA	ACCAAGGARCARTTCAAG	*mmoX*	6
mmoXB	TGGCACTCRTARCGCTC	*mmoX*	6

The attached communities were sampled through the use of substrate columns. These columns consisted of crushed basalt chips (4 to 7 mm sieve size) housed in a 76-cm-long nylon mesh tubing (inside diameter, 2 cm; mesh size, 0.25 cm; InterNet, Inc., Minneapolis, Minn.). Prior to use, the columns were rinsed with deionized water to remove fine particles and then autoclaved (60 min) on three successive days. The sterile columns were then suspended in the well to a depth of 80 m for 6 months. At the end of this period, the columns were removed and the attached microbial mass was harvested.

Metagenomic DNA isolation. The UltraClean Mega Prep soil DNA kit (MoBio Laboratories, Inc., Solana Beach, Calif.) was used to harvest the attached community metagenomic DNA from the basalt chips. Four 50-ml polypropylene tubes were filled with 40 g of the substrates and secured horizontally in a rotating incubator (Innova 4300; New Brunswick Scientific Co., Inc., Edison, N.J.) at 200 rpm at 60°C for 30 min to remove bacterial biomass from the basalt substrates. Metagenomic DNA was recovered as per the manufacturer's protocol. After recovery, the extractions were pooled and concentrated by using a Microcon YM-30 ultrafiltration centrifugal device (Millipore, Billerica, Mass.).

Free-living cells were recovered by filtering 50 ml of the concentrated material through a 25-mm (0.20-μm)-pore-size Nuclepore polycarbonate membrane filter (Whatman, Inc., Clifton, N.J.). DNA was then extracted from the filtered material by use of the UltraClean soil DNA isolation kit from MoBio. Briefly, the filter was placed into the bead solution and vortexed on high for 10 s to disperse the filtered material. The tubes were then placed in a 65°C water bath for 30 min. The purification of the lysed material was performed as per the manufacturer's protocol. This method was performed on three additional aliquots so that a total concentrate volume of 200 ml was processed. The resulting DNA from each extraction was pooled and stored at −20°C. DNA recovered from the substrate-attached bacteria by the modified MoBio Mega Prep method resulted in 140 μl of DNA at a concentration of 24 ng μl^{-1} (A_{260}/A_{280} = 1.5).

Whole-genome amplification. Purified DNA from both communities was amplified by using the Repli-*g* whole-genome amplification (WGA) kit (Molecular Staging, Inc., New Haven, Conn.) to produce two additional community archives, W9 and W10. Briefly, the kit utilizes ϕ29 polymerase in conjunction with exonuclease-resistant random hexamer primers to isothermally amplify DNA 10^4- to 10^6-fold with minimal bias (26). Based on multiple-displacement amplification technology, exponential amplification occurs through a hyperbranching mechanism (12, 25). Long products, averaging 12 kb in length, are obtained due to the high processivity rate of >70,000 nucleotides incorporated for each primer binding event (8). ϕ29 also exhibits the highest fidelity rate of any known polymerase, with an error rate of only 1 in 10^6 to 10^7 nucleotides incorporated (14).

For each sample, 10 ng of genomic DNA was chemically denatured by being mixed with 2.5 μl of denaturation solution (40 mM KOH, 1 mM EDTA [pH 8.0]) and incubated at room temperature for 3 min. The reaction mixture was neutralized by the addition of 5 μl of neutralization solution (a 1:10 dilution of supplied solution B in distilled water). Amplification reactions were performed in 50-μl volumes containing 12.5 μl of 4× WGA PCR mix (as supplied by the manufacturer), 0.5 μl of DNA polymerase mix, 27 μl of sterile H_2O, and 10 μl of denatured sample. The reaction mixtures were then incubated at 30°C for 16 h. Amplification of the free-living DNA was only successful after two washes with distilled water followed by concentration with a Microcon filter device. Amplified products were visualized by using 1% agarose gel electrophoresis stained with ethidium bromide. The amplified DNA was quantified by densitometry using Kodak 1D software (Kodak, Rochester, N.Y.).

Functional gene PCR amplification. Three sets of PCR primers were used in this study to elucidate the oxygenase genes associated with methane- and ammonia-oxidizing bacteria from each community (attached and free-living) and WGA material (W9 and W10) (Table 1). To date, no single primer set has been

demonstrated to amplify *mmoX* genes from all sMMO-producing bacteria.

Primers that targeted the *mmoX* gene and the *pmoA* gene specifically were used, as was a set that allowed the corecovery of *amoA* gene fragments along with *pmoA*. All PCR products were verified on 1% agarose gels stained with ethidium bromide and visualized by using UV light. Images were captured and recorded by using Kodak 1D software. Reagents and enzymes were purchased from Fisher Scientific (Hampton, N.H.) unless otherwise specified. Oligonucleotides were synthesized by Integrated DNA Technologies, Inc. (Coralville, Iowa) with standard desalting.

pMMO diversity within the communities was assessed through the use of primers designed by Holmes et al. (19) and Costello and Lidstrom (11) (Table 1). Both groups utilized the same forward primer, A189, in conjunction with one of two reverse primers. The first reverse primer, A682, was designed to target a highly conserved region in both the *pmoA* and *amoA* genes. These genes are believed to encode the active site containing the 27-kDa polypeptide subunit of both the MMO and the AMO. The A189-A682 primer set has been shown previously to amplify a 525-bp internal section of both *pmoA* and *amoA* (19). The second reverse primer, mb661, has been shown to preferentially amplify only an internal section of approximately 470 bp from the *pmoA* gene (11). PCR amplifications with both primer sets were performed in 50-μl volumes containing a 200 μM concentration of each deoxynucleoside triphosphate, 1× PCR buffer, 10 μg of bovine serum albumin, 1.2 mM $MgCl_2$, 10 pmol of each forward and reverse primer, 2.5 U of *Taq* polymerase, and 20 ng of DNA template. Reaction mixtures were subjected to initial denaturation for 5 min at 94°C followed by 30 cycles of 57°C for 1 min, 72°C for 1 min, and 94°C for 1 min. The final step consisted of a 57°C annealing for 1 min, followed by a 10-min elongation at 72°C.

The primer set mmoXA and mmoXB was used to amplify an approximately 1,230-bp fragment from the *mmoX* gene as described by Auman et al. (6). This primer set is designed to target the region of *mmoX* that encodes the conserved fragment of the α subunit of the hydroxylase component of sMMO. PCR amplifications were performed in 50-μl volumes containing 200 μM concentrations of each deoxynucleoside triphosphate, 1× PCR buffer, 10 μg of bovine serum albumin, 1.2 mM $MgCl_2$, 10 pmol of each forward and reverse primer, 2.5 U *Taq* polymerase, and 20 ng of DNA template. Reaction mixtures were subjected to an initial denaturation of 5 min at 94°C followed by 30 cycles of 60°C for 1 min, 72°C for 1 min, and 94°C for 1 min. The final step consisted of a 60°C annealing for 1 min, followed by 10 min of elongation at 72°C.

A PCR on a metagenomic DNA template isolated from an undefined groundwater methanotroph enrichment which had been grown on methane was used as a positive control for each primer set.

Following amplification, all PCR products were purified by a Wizard spin column PCR cleanup (Promega, Madison, Wis.). Some products required additional purification by agarose gel electrophoresis in order for clones with the proper insert size to be obtained. This procedure was done by loading approximately 1 μg of PCR product into 1% SeaPlaque GTG LMP agarose, followed by extraction via Qiaex II gel extraction (QIAGEN, Valencia, Calif.).

Archives were coded based on the well numbers used in the sampling (9 and 10 for free-living and attached communities, respectively) and the primer sets used in the PCR amplification. For instance, the archive produced from free-living cells isolated from well ANP-9 and amplified with the PCR set A189-A682 was coded as 9-682. Archives produced from whole-genome-amplified DNA were designated with a "W" preceding the well identifier.

Functional gene fragment library production. PCR amplicons for each sample were cloned by using the TOPO TA cloning kit for sequencing (version K) (Invitrogen Corporation, Carlsbad, Calif.). Briefly, 4 μl of each fresh PCR product was incubated with 1 μl of salt solution (1.2 mM NaCl, 0.06 M $MgCl_2$) and 1 μl of pCR4-TOPO vector. Electrocompetent Top10 *Escherichia coli* cells were electroporated at 2.4 kV in 0.2-cm cuvettes with 2 μl of the cloning reaction mixture. Positive transformants were selected by spread plating onto Luria-Bertani agar plates with either kanamycin (50 μg ml^{-1}) or ampicillin (60 μg ml^{-1}) used as the selective agent and incubated overnight at 37°C. Kanamycin was used when direct PCR amplicons were cloned, whereas ampicillin was used when the amplicon was purified by gel electrophoresis.

Ninety-six colonies were isolated for each library archive and inoculated into 96-well plates containing 200 μl of Luria-Bertani broth and 50 μg of kanamycin ml^{-1} amended with 1× Haugness buffer (4% glycerol, 3.6 mM K_2HPO_4, 1.3 mM KH_2PO_4, 2 mM trisodium citrate, 1 mM $MgSO_4$, pH 7.5) as a cryoprotectant. These plates were then incubated in a GeneMachines HiGro microtiter-plate orbital shaker (Genomic Solutions, Ann Arbor, Mich.) for 18 h at 420 rpm and 37°C with an airflow rate of 4 standard liters per min per chamber. Ten wells were randomly sampled for insert verification by PCR using vector-specific flanking primers (M13f and M13r supplied with the kit). Amplicons were visualized on 1% agarose gels prestained with ethidium bromide. After verification, plates were sealed with aluminum tape and stored at −80°C for future use.

Restriction fragment length polymorphism (RFLP) analysis. Plasmid DNA was isolated from the 96-well cultures by using the Montage Plasmid $Miniprep_{96}$ kit (Millipore). Cultures were prepared by inoculating one round-bottom, 96-well cell culture block (supplied with the kit) containing 1 ml of Terrific broth with 50 μg of kanamycin ml^{-1} per well for each archive. Blocks were incubated in the HiGro incubator for 24 h under the conditions described above. Cells were harvested by centrifugation at 1,500 × *g* for 5 min with subsequent removal of supernatant by decanting. The blocks were sealed and pellets were frozen at −20°C prior to processing.

Plasmid DNA was isolated from the cell pellets as per the manufacturer's full lysate protocol. Five wells were selected from each plasmid preparation and DNA quantified by measuring the absorbance at 260 nm.

RFLP patterns were produced for each clone by the digestion of 1 μg of plasmid DNA with 5 U each of MspI and HinPII (New England Biolabs, Beverly, Mass.) in a final reaction mixture volume of 50 μl. Reaction mixtures were incubated at 37°C for 2 h, followed by enzyme inactivation at 65°C for 10 min. The manual grouping of fragmentation patterns was performed by analyzing 10 μl of each digest on 4% NuSieve 3:1 agarose (Cambrex Bio Science, Baltimore, Md.). Gels were stained with ethidium bromide and rinsed with deionized H_2O prior to imaging. One representative from each RFLP group was chosen for further analysis.

Sequencing and phylogenetic analysis. Sequencing of insert DNA was performed on an ABI Prism model 3730 sequencer (Applied Biosystems, Inc., Foster City, Calif.) by using BigDye version 3 chemistry. Contiguous sequences were produced by using ContigExpress software of the Vector NTI suite, version 9. Nucleotide-nucleotide BLAST (blastn) was used to search GenBank for nearest relative sequences. BLAST results and representatives for each archive were aligned by using AlignX software from the Vector NTI suite, version 9.

Edited alignments were evaluated with the maximum likelihood method by using the PAUP package (version 4.0b10; Sinauer Associates, Inc., Sunderland, Mass.). Evolutionary distance calculations were generated by using the model best describing the data for each tree as determined using the DT-ModSel program (38). Confidence estimates for the nodes within phylogenetic trees were performed by bootstrap analysis (100 replicates). Trees were visualized with TreeView software, version 1.6.6 (43).

RESULTS

Sampling. Free-living cell collection produced 3 liters of concentrate at 4.2×10^6 cells ml^{-1} (1.26×10^{10} cells in total), as determined by direct microscopic counts of 4′,6-diamidino-2-phenylindole (DAPI)-stained cells. After taking into account a 1:4,433 dilution factor, these results suggest a concentration of approximately 10^3 cells ml^{-1} in the original ground water. This number agrees with previous cell density estimates from this same well (40). Biomass from the substrate columns was not quantified.

Metagenomic DNA isolated from sampling of free-living cells by the heat-lysis method used in conjunction with the MoBio kit produced a total of 160 μl of DNA at 12.5 ng μl^{-1} ($A_{260}/A_{280} = 1.61$). The DNA was ~25 kb in size.

Whole-genome amplification. In an effort to assess the applicability of using whole-genome amplification as a means to generate additional metagenomic template DNA for analyses, archives generated from amplified and unamplified templates were compared. WGA samples resulted in quantities of 550 ng μl^{-1} and 800 ng μl^{-1} (quantified by densitometry) for the free-living and attached communities, respectively. These concentrations represent nearly 10^4-fold amplifications. Both amplified samples migrated between 20 and 25 kb when analyzed by agarose gel electrophoresis.

PCR amplification. DNA extracted from the free-living and attached communities resulted in PCR amplicons of the expected size for a given primer pair. The only exception was the

absence of a PCR product from DNA isolated from the attached community with the *mmoX*-specific primers. These primers were also unsuccessful at amplifying the targeted gene fragment from WGA attached community DNA. In some instances, the products did not resolve as single, isolated bands even after PCR optimization.

Gene fragment library production. It was found initially that high transformation rates could be obtained, although a large percentage (30 to 50%) of those clones contained nonspecific inserts. In order to increase the percentage of clones containing the proper insert size, it was necessary to gel purify the PCR product prior to cloning. This was the case even in the absence of other visible contaminating DNA, such as multiple bands or primer dimers (data not shown). After gel purification, the percentage of clones not containing inserts of the proper size dropped well below 10%.

In general, DNA from all communities transformed at a very low efficiency compared to positive controls supplied with the kit, though some efficiencies were far lower than others. The reason for these differences remains unclear, though we speculate that it may be due to the nature of the oligonucleotide primers used in the PCR. In most cases, multiple cloning and transformation reactions were necessary to produce 96 colonies for the archives.

RFLP and phylogenetic analysis. A total of 96 clones were screened for each of the 10 functional gene archives. Initially, many RFLP patterns were observed in all archives, and a majority of those groups contained only a single representative. Subsequent sequencing revealed that many of these patterns were a result of false-positive clones that did not contain the proper insert. These RFLP groups were discarded from the archives. Figure 1 shows a representative gel containing multiple RFLP patterns from 9-661, 9-682, W9-661, W9-682, and W10-682.

Archives produced with primer set A189-A682 resulted in a combined total of 327 clones. From these clones, only type II methanotroph and AOB sequences that corresponded to α-*Proteobacteria* and β-*Proteobacteria*, respectively, were found (Fig. 2). Archive compositions of whole-genome amplified and unamplified communities were similar except that group I uncultured (*pmoA*) sequences were obtained with less frequency in the WGA free-living community archive than in the unamplified archive, while *Nitrosospira* spp. (*amoA*) were represented to a much greater degree in the WGA-derived archive than in the unamplified archive. Interestingly, this difference was not observed for the attached community. A comparison of the free-living and attached communities showed differences between the two communities. Group II uncultured (*pmoA*) sequences were more prevalent in the attached community than in the free-living community. Furthermore, representative sequences from group I and III uncultured (*pmoA*) bacteria and *Nitrosospira* were not found in the attached community archive (Table 2).

Archive 9-682 produced a total of 83 clones from which 13 unique RFLP patterns were obtained (Fig. 2). Just over half (59%) of the clones were found to contain oxygenase genes that group within α-*Proteobacteria*, while the remainder (41%) were from subclass β-*Proteobacteria*. BLAST results for groups 1, 3 to 7, 11, 13, 16, and 19 revealed sequences closely (95 to 98%) related to those of previously identified *pmoA*/*amoA*-containing bacteria. However, groups 2, 9, and 17 showed only 88% sequence identity to the oxygenase genes of uncultured bacterium M84-P36. These three groups branched separately from M84-P36 and formed a distinctive cluster with W9-682 7 (Fig. 2).

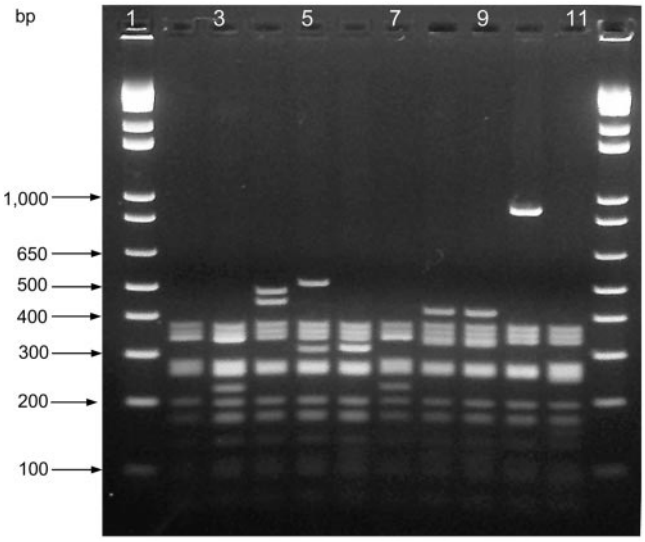


FIG. 1. Representative RFLP analysis. Plasmid DNA isolated from clones produced in the functional gene archives was digested with HinPII. Two hundred fifty nanograms of each digest was separated by using 2% LMP GTG agarose run at 5 V cm^{-1} for 1.75 h in 1% Tris-acetate-EDTA. Lanes 1 and 12 contain 500 ng of 1-kb Plus DNA molecular weight marker. Lanes 2 and 3, 9-661; lanes 4 and 5, 9-682; lanes 6 and 7, W9-661; lanes 8 and 9, W9-682; lanes 10 and 11, W10-682. The gel was stained with ethidium bromide. Images were captured and recorded by using Kodak 1D software and UV transillumination.

W9-682 produced a total of 91 clones that resulted in 10 unique RFLP patterns, of which 14% were α-*Proteobacteria* and 86% were β-*Proteobacteria* (Fig. 2). All groups, with the exception of group 7, yielded high sequence identity (94 to 99%) to previously identified *pmoA*/*amoA*-containing bacteria similar to those in 9-682. Group 7, like groups 2, 9, and 17 of 9-682, had a highest identity value of 88% for the uncultured bacterium M84-P36. Group 11 formed a distinct branch within the tree.

Archives 10-682 and W10-682 were both very similar in composition and showed less oxygenase diversity than the free-living community archives, as only five RFLP patterns were seen for 10-682 and six patterns were seen for W10-682 (Fig. 2). Both archives contained a large majority (89 and 97%, respectively) of β-*Proteobacteria*, of which nearly all showed high levels of similarity to oxygenase genes of the uncultured bacterium gp22 and are distantly related to the AOB *Nitrosospira* (43 of 45 nucleotides were identical). Two of the groups of α-*Proteobacteria*, 10-682 4 and W10-682 10, clustered near the MOB *Methylocystis* sp. 42/22, while the third group of W10-682 15 clustered near other uncultured α-*Proteobacteria*.

Archives produced with primer set A189-mb661 resulted in a combined total of 317 clones. An overall high percentage of α-*Proteobacteria* was found in each of these archives. Archive 9-661 and W9-661 contained 97 and 77% *Methylocystis*, respec-

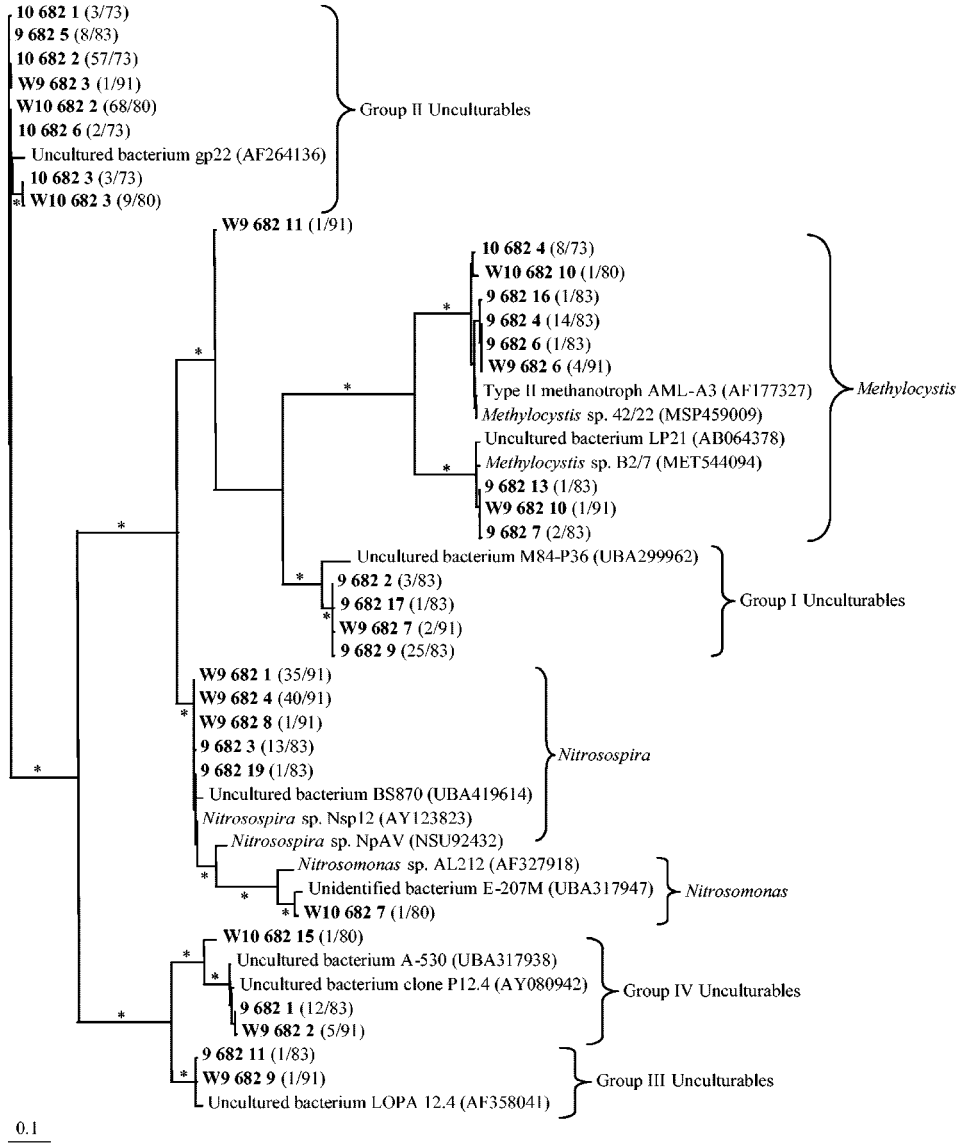


FIG. 2. Phylogenetic relationships among partial *amoA*/*pmoA* gene sequences produced from archives of unamplified and WGA free-living and attached communities. Gene fragments were obtained by using the primer set A189-A682 (540 base positions were considered). The bar represents 10% sequence divergence as determined by the length of the horizontal lines connecting any two species. The phylogenetic tree was constructed by using PAUP based on maximum likelihood analysis using the K80+G correction model. Bootstrap values greater than 75% are represented by asterisks. Unique clones obtained from the archives are indicated by boldface type. Each clone is named in the following manner: the first number (9 or 10 in bold) refers to the well number, the second number (e.g., 682 in bold) refers to the reverse primer that was used to obtain this sequence, the third number (in bold) is the RFLP group number, and the numbers in parentheses refer to the fraction of clones represented by this phylotype.

tively, and both 10-661 and W10-661 were made up of 100% *Methylocystis*-like sequences (Table 3).

Archives 9-661 and W9-661 were very similar in composition, with a high percentage of each corresponding to *Methylocystis* sp. (Fig. 3). Seven RFLP patterns were seen in archive 9-661, whereas W9-661 resulted in 13 patterns. Four of the groups, 9-661 9, W9-661 5, W9-661 11, and W9-661 13, branched with the type I methanotroph *Methylomonas* sp., though they formed a divergent cluster. Another type I methanotroph *Methylobacter* sp. was distantly related (81% sequence identity) to the groups 9-661 3 and W9-661 12. The nearest identity of these two groups is to the uncultured bacterium LP20 (89%), though the sequences were divergent enough to branch separately.

Archives 10-661 and W10-661 each produced six unique RFLP patterns and were similar in composition (Fig. 3). All of the clones screened in each archive were highly related to oxygenase genes of *Methylocystis*, with >98% sequence identities in most cases and >95% sequence identities in all cases. These clones clustered very closely to the *Methylocystis* sp. used as the reference in the phylogenetic tree. No type I methanotrophs or AOB were found.

TABLE 2. Diversity of *pmoA* and *amoA* functional genes obtained in archives generated by primer set A189-A682 for free-living (9-682 and W9-682) and attached (10-682 and W10-682) microbial communities within the ESRPA

Group or species (relevant gene)	% Archive composition			
	Free-living		Attached	
	9-682	W9-682	10-682	W10-682
Group I uncultured (*pmoA*)	35	2	0	0
Group II uncultured (*amoA*)	10	1	89	96
Group III uncultured (*pmoA*)	1	1	0	0
Group IV uncultured (*amoA*)	14	6	0	1
Nitrosospira sp. (*amoA*)	17	84	0	0
Methylocystis sp. (*pmoA*)	23	6	11	1
Nitrosomonas sp. (*amoA*)	0	0	0	1
Total	100	100	100	99

Archives produced from primer set mmoXA-mmoXB resulted in a total of 143 clones. Of these clones, 12 RFLP patterns were produced from 9-mmoX, and 7 patterns were produced from W9-mmoX. Overall, both archives resulted in similar compositions, with *Methylocystis*-like oxygenase sequences resulting in 95 and 100% of the total clones for 9-mmoX and W9-mmoX, respectively (Table 4).

Phylogenetic analysis of the two archives revealed three distinct clusters, two closely related to *Methylocystis* sp. and the other to *Methylomonas* sp. Type I methanotrophs found only in 9-mmoX (groups 5, 8, and 11) were related to *Methylomonas* sp. LW15 (>91% sequence identity). These three groups collectively account for 5% of the overall archive. Type II methanotrophs represented in 9-mmoX include 87% *Methylocystis* sp. IMET 10486 (groups 1, 2, 4, and 14), 7% *Methylocystis* sp. 51 (groups 3, 13, 15, and 16), and 1% *Methylocystis* sp. SE12 (group 12). Type II methanotrophs represented in W9-mmoX include 68% *Methylocystis* sp. SE12 (groups 1, 5 and 7), 30% *Methylocystis* sp. IMET 10486 (groups 6, 10, and 12), and 2% *Methylocystis* sp. 51 (group 14) (Fig. 4).

DISCUSSION

Sampling. It has been reported that most bacteria in an aquifer are attached to the aquifer solids, and only a small fraction exist as free-living cells (18, 29). The sampling of attached biomass from aquifer boreholes is usually accomplished through the collection of core samples. This method can be expensive, as well as challenging, depending on the aquifer to be sampled. In this study, we chose to use porous columns filled with material simulating the existing fractured basalt of the aquifer and incubated in a borehole in the aquifer.

TABLE 3. Diversity of the *pmoA* functional gene obtained in archives generated by primer set A189-mb661 for free-living (9-661 and W9-661) and attached (10-661 and W10-661) microbial communities within the ESRPA

Species	% Archive composition			
	Free-living		Attached	
	9-661	W9-661	10-661	W10-661
Methylocystis sp.	97	77	100	100
Methylomonas sp.	1	17	0	0
Methylobacter sp.	2	6	0	0
Total	100	100	100	100

The use of artificial substrata incubated in wells to sample attached communities close to the study site described herein has been described recently by Lehman et al. (29). The study by Lehman et al. found that the use of dialysis chambers containing basalt (similar to the substrate columns used here) resulted in the recovery of a greater range of microbial diversity compared to that from authentic core samples. Therefore, boreholes may not be ideal for collecting representative samples of microbes colonizing aquifer solids. Free-living communities sampled by filtration of groundwater are also subject to biases. Due to the large volume of ground water used in this study (13,300 liters), it is likely that many different microbial niches were sampled, not all necessarily corresponding to the free-living community. Sediment particles were present in the filtrate, and bacteria colonizing these sediment layers would have been extracted along with free-living cells. It has been observed that bacteria display a dynamic equilibrium between attached and free-living phases (51). These bacteria form biofilms as part of their normal life cycle that culminates with the release of free-living cells capable of colonizing new habitats. Thus, it is difficult to draw a distinct line between attached and free-living communities. Though the communities may differ in overall composition, it is not surprising that we found no distinct disparity between the attached and free-living communities sampled. It is likely that free-living cells present in the free-living samples are also represented, to some extent, in the attached samples.

Whole-genome amplification. WGA technology is based on the unbiased amplification of limited DNA samples (21, 26). This technology was able to amplify our starting DNA to produce a product equal in size to the starting template when visualized by agarose gel electrophoresis (data not shown). PCR of WGA products also resulted in amplicons equal in size to unamplified samples. The differences noted between archives were most likely due not to the WGA process but rather to bias introduced by the PCR. Since multiple PCRs were not pooled, bias generated through a single PCR amplification may have skewed the resulting archives.

Functional gene library production and phylogenetic analysis. The primer sets A189-A682, A189-mb661, and mmoXA-mmoXB were used successfully as functional gene probes in the assessment of MOB and AOB in most of the communities tested. It is unclear why no amplification was seen from the attached and WGA attached DNA with the mmoXA-mmoXB primer set. Sequences were amplified from both attached and WGA attached communities by using the primer sets A189-A682 and A189/mb661, and these were closely related to sequences of the sMMO-producing MOB *Methylocystis*.

An explanation for our observations may lie in the distinct difference in the species of *Methylocystis* found in free-living versus attached communities. None of the sequences amplified from the free-living samples with A189-A682 and A189-mb661 was found in the mmoXA-mmoXB archives and vice versa. It appears that the three primer sets used exhibit specificity towards different *Methylocystis* spp. This fact supports the conclusion that the methanotroph species targeted by mmoXA-

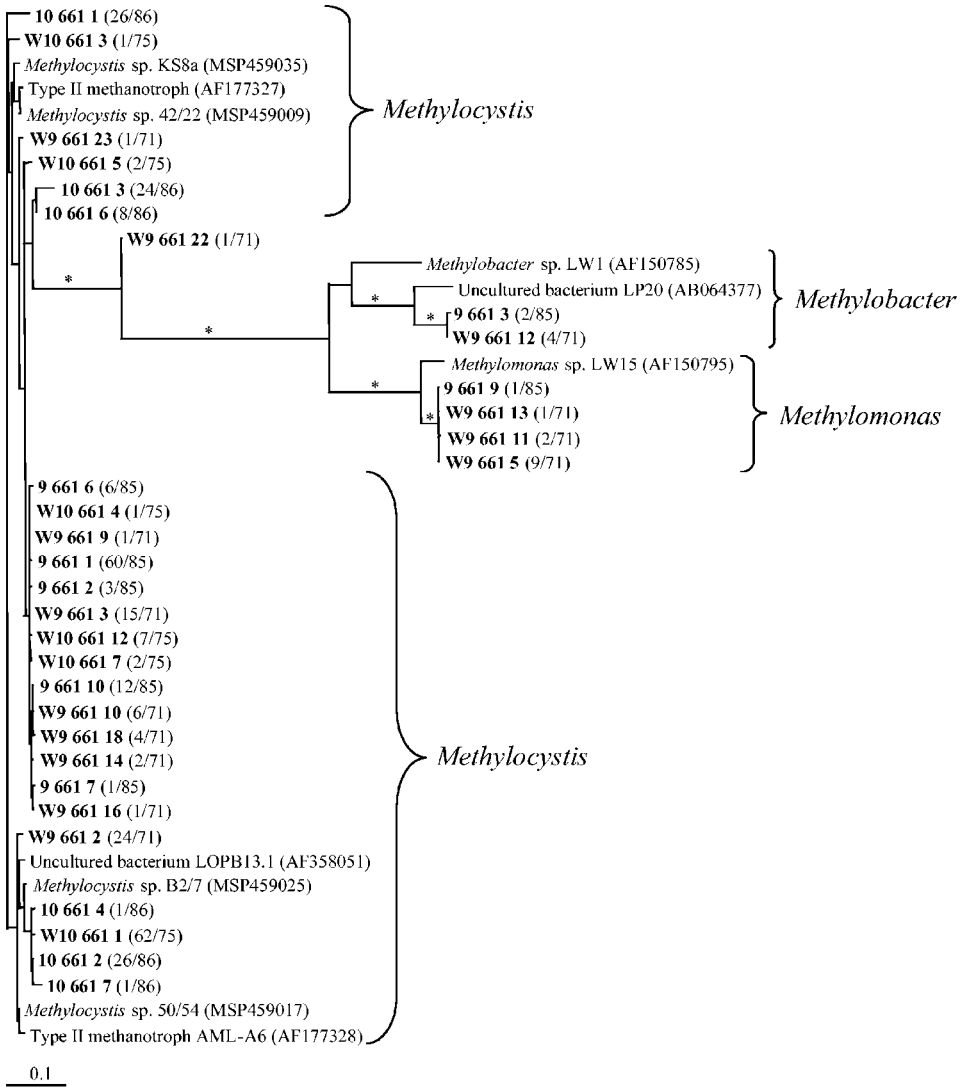


FIG. 3. Phylogenetic relationships among partial *pmoA* gene sequences produced from archives of unamplified and WGA free-living and attached communities. Gene fragments were obtained by using the primer set A189-mb661 (550 base positions were considered). The bar represents 10% sequence divergence as determined by the length of the horizontal lines connecting any two species. The phylogenetic tree was constructed by using PAUP based on maximum likelihood analysis using the HKY+G correction model. Bootstrap values greater than 75% are represented by asterisks. Unique clones obtained from the archives are indicated by boldface type. Clones are named according to the manner described in the legend to Fig. 2.

mmoXB were not present in the attached samples in sufficient numbers to be detected.

The number of available *mmoX* sequences on which to base primer design is small. The primer set *mmoX* f882-*mmoX* r1403, widely used for *mmoX* amplification, is based on sMMO gene clusters from only two methanotrophs (34). The mmoXA-mmoXB primer set, which was designed 5 years after the introduction of *mmoX* f882-*mmoX* r1403, is based on only six known full *mmoX* sequences (6). Deduced amino acid alignments show a high percentage of conservation from sequences amplified by Auman et al. (6) with these primers. This result could be due to the fact that the primer set is specific for only highly unique genes. Another study done by Auman and Lidstrom failed to amplify any type II *mmoX* genes from their samples, even though type II-specific 16S rRNA PCR showed these bacteria to be present (5). Since we observed a similar result in a different environment, it may be that genes found to be conserved in cultivable organisms may not be sufficiently similar to genes in the environment, as noted by Hanson and

TABLE 4. Diversity of the *mmoX* functional gene obtained in archives generated for free-living communities within the ESRPA

Species	% Archive composition	
	9-mmoX	W9-mmoX
Methylocystis sp.	95	100
Methylomonas sp.	5	0
Total	100	100

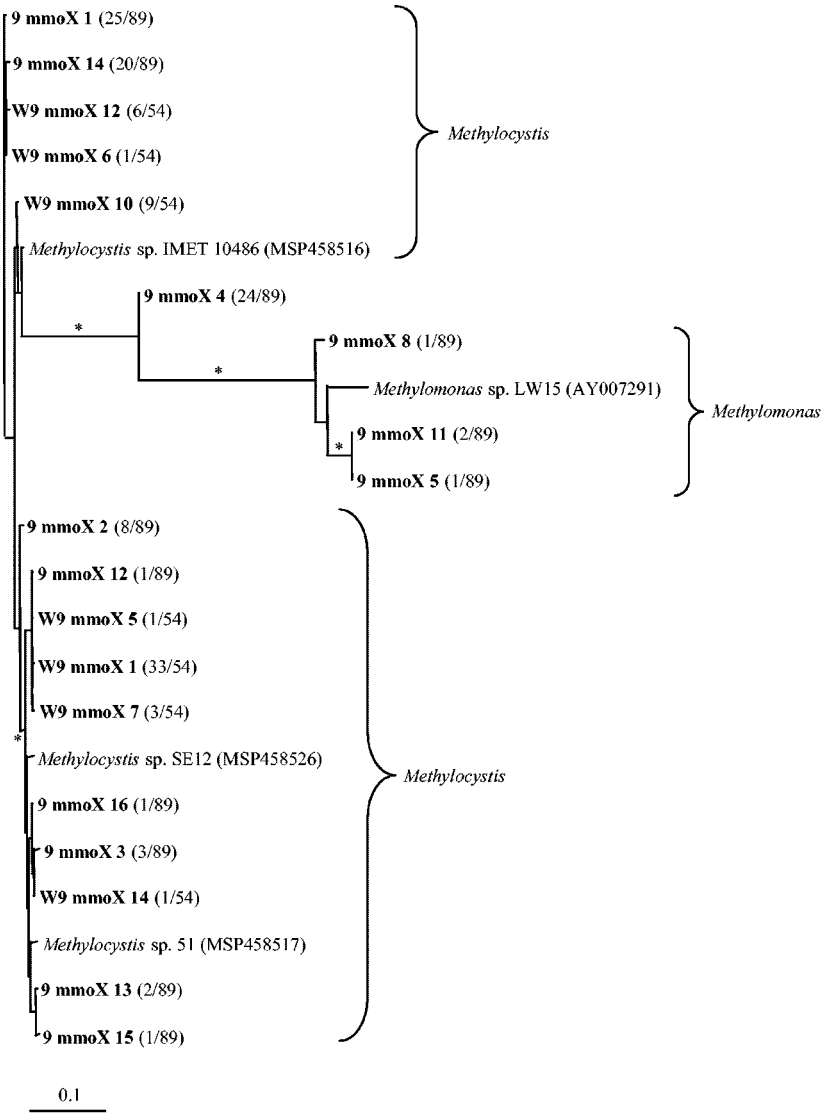


FIG. 4. Phylogenetic relationships among partial *mmoX* gene sequences produced from archives of unamplified and WGA free-living communities. Gene fragments were obtained by using the primer set mmoXA-mmoXB (1,271 base positions were considered). The bar represents 10% sequence divergence as determined by the length of the horizontal lines connecting any two species. The phylogenetic tree was constructed by using PAUP based on maximum likelihood analysis using the TrN+G correction model. Bootstrap values greater than 75% are represented by asterisks. Unique clones obtained from the archives are indicated by boldface type. Clones are named according to the manner described in the legend to Fig. 2.

Hanson (17), and therefore must be considered when interpreting data generated by such primer sets. To date, no single primer set has been demonstrated to amplify *mmoX* genes from all sMMO-producing bacteria.

As expected, the primer set A189-A682 amplified sequences corresponding to both *pmoA* and *amoA* from all four communities. The abundance of *amoA*-related sequences compared to *pmoA* found in all four archives except 9-682 suggests that nitrifying bacteria might dominate the samples. Of particular interest is the presence of the large subset of clones that cluster with the uncultured bacterium M84-P36. These clones make up a total of 35% of the archive and are only distantly related to other uncharacterized *pmoA* clones (20), suggesting the presence of a potentially significant novel group of *pmoA*-carrying methanotrophs. The proportion of sequences in clone libraries, however, cannot be used to accurately predict the relative abundance of the AOB or the novel *pmoA* group in the natural communities, due to biases that are inherent in PCR amplification and the number of sequences analyzed. Such proportions, however, are useful for comparisons of archive data within a given data set.

Characterization of the communities with *pmoA*-specific primers (A189-mb661) suggested that bacteria related to the type II methanotroph *Methylocystis* dominate free-living as well as attached methanotroph communities. Newby et al. obtained similar results in their study of the methanotroph communities from the ESRPA (40). However, their study sampled a much smaller volume of groundwater and characterized only free-

living cells. The free-living communities (amplified and unamplified) also showed the presence of type I methanotrophs related to *Methylomonas* sp. LW15, a known sMMO-containing bacterium (6). The small percentage of clones from archives 9-661 and W9-661 found to be related to the genus *Methylobacter*, which has not been shown to produce sMMO, suggests that a broad range of niches may have been sampled through the collection of free-living cells. Again, the Newby et al. study documented the presence of these type I methanotrophs by using a different molecular approach (40). A direct comparison of the gene fragments could not be made, due to differences in the targeted areas of the genes in question.

It has been proposed that growth of type II methanotrophs is favored over growth of type I methanotrophs under low-oxygen, high-methane conditions (17). The analysis of groundwater sampled from well ANP-9 showed a significant concentration of methane but also revealed an oxic environment (40). It is expected that an environment such as this should favor the growth of type I methanotrophs. The presence of copper, however, is another key factor responsible for regulating the growth of type I and type II methanotrophs. Copper is necessary for pMMO activity and plays a key role in the regulation of pMMO and sMMO expression (30). Under low-copper conditions, bacteria with the ability to produce sMMO are able to outcompete those able to produce only pMMO (17, 39). This mechanism is due to the fact that the active pMMO enzyme contains approximately 12 to 15 copper atoms mol^{-1} (41). While the exact nature of these Cu^{2+} ions in the function of pMMO is not clear, they are postulated to play a role in the active site of the protein (41, 44). Samples of groundwater from ANP-9 show that the copper levels are below the detection limit of 1.0 $\mu g\ ml^{-1}$ (23). Taking into account the conditions present in the wells sampled (the presence of dissolved methane and low copper concentrations), the community composition observed in this study is not surprising.

Our results, obtained through the molecular assessment of oxygenase genes, demonstrate that in both the attached and free-living communities, the majority of the methanotrophic populations present show a high degree of similarity to other type II methanotrophs. The natural attenuation of TCE observed at the TAN site may be due in part to the sMMO-producing communities described here. However, we did not attempt to enumerate methanotrophs from either free-living or attached communities. Because of this fact, the organisms represented in this study constitute an undetermined percentage of the overall bacterial population. Recently, Lehman et al. found that very few MOB or AOB could be recovered from authentic core or groundwater samples taken from within the TCE plume (29). The results from their study may reflect the fact that the enumerations were based on most-probable number assays performed with medium selective for type I methanotrophs and would not have detected the type II populations described here.

The presence of sMMO-producing bacteria alone does not account for the rate of TCE natural attenuation observed at the TAN site. It was our goal to provide evidence that the in situ bacterial populations within the pristine aquifer have the potential for TCE cometabolism. Future work to quantify type II methanotrophs specifically and monitor the production of sMMO is necessary in order to correlate the presence of these bacteria with the TCE attenuation rate observed at the TAN site within the contaminated plume.

ACKNOWLEDGMENTS

This project was funded by Department of Energy Environmental Management Science Program grant no. DE-FG07-02ER63500, National Institutes of Health grant no. 1R21 ES012814-01, and Inland Northwest Research Alliance Task Order no. UI002.

REFERENCES

1. **Alvarez-Cohen, L., P. L. McCarty, E. Boulygina, G. Brusseau, and R. S. Hanson.** 1992. Cometabolic biotransformation of trichloroethylene and chloroform by a bacterial consortium grown with methane. Appl. Environ. Microbiol. **58:**1886–1893.
2. **Alvarez-Cohen, L., and G. E. Speitel, Jr.** 2001. Kinetics of aerobic cometabolism of chlorinated solvents. Biodegradation **12:**105–126.
3. **Anderson, J. E., and P. L. McCarty.** 1997. Transformation yields of chlorinated ethenes by a methanotrophic mixed culture expressing particulate methane monooxygenase. Appl. Environ. Microbiol. **63:**687–693.
4. **Arp, D. J., C. M. Yeager, and M. R. Hyman.** 2001. Molecular and cellular fundamentals of aerobic cometabolism of trichloroethylene. Biodegradation **12:**81–103.
5. **Auman, A. J., and M. E. Lidstrom.** 2002. Analysis of sMMO-containing type I methanotrophs in Lake Washington sediment. Environ. Microbiol. **4:**517–524.
6. **Auman, A. J., S. Stolyar, A. M. Costello, and M. E. Lidstrom.** 2000. Molecular characterization of methanotrophic isolates from freshwater lake sediment. Appl. Environ. Microbiol. **66:**5259–5266.
7. **Barnes, J. M., C. Rae, V. S. Thompson, M. R. Walton, D. T. Maiers, and K. S. Sorenson.** 2000. Aerobic biodegradation laboratory studies at Test Area North, Operable Unit 1-07B (INEEL/EXT-99-00736). U.S. Department of Energy, Idaho Operations Office, Idaho Falls, Idaho.
8. **Blanco, L., A. Bernad, J. M. Lazaro, G. Martin, C. Garmendia, and M. Salas.** 1989. Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. J. Biol. Chem. **264:**8935–8940.
9. **Brusseau, G. A., H. C. Tsien, R. S. Hanson, and L. P. Wackett.** 1990. Optimization of trichloroethylene oxidation by methanotrophs and the use of a colorimetric assay to detect soluble methane monooxygenase activity. Biodegradation **1:**19–29.
10. **Burrows, K. J., A. Cornish, D. Scott, and I. J. Higgins.** 1984. Substrate specificities of the soluble and particulate methane monooxygenases of *Methylosinus trichosporium* OB3b. J. Gen. Microbiol. **130:**3327–3333.
11. **Costello, A. M., and M. E. Lidstrom.** 1999. Molecular characterization of functional and phylogenetic genes from natural populations of methanotrophs in lake sediments. Appl. Environ. Microbiol. **65:**5066–5074.
12. **Dean, F. B., S. Hosono, L. Fang, X. Wu, A. F. Faruqi, P. Bray-Ward, Z. Sun, Q. Zong, Y. Du, J. Du, M. Driscoll, W. Song, S. F. Kingsmore, M. Egholm, and R. S. Lasken.** 2002. Comprehensive human genome amplification using multiple displacement amplification. Proc. Natl. Acad. Sci. USA **99:**5261–5266.
13. **Dedysh, S. N., N. S. Panikov, W. Liesack, R. Grosskopf, J. Zhou, and J. M. Tiedje.** 1998. Isolation of acidophilic methane-oxidizing bacteria from northern peat wetlands. Science **282:**281–284.
14. **Esteban, J. A., M. Salas, and L. Blanco.** 1993. Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. J. Biol. Chem. **268:**2719–2726.
15. **Federal Register.** 1989. National primary and secondary drinking water regulations. Fed. Regist. **54:**48184–48189.
16. **Federal Register.** 1989. National primary and secondary drinking water regulations. Fed. Regist. **54:**22062–22160.
17. **Hanson, R. S., and T. E. Hanson.** 1996. Methanotrophic bacteria. Microbiol. Rev. **60:**439–471.
18. **Holm, P. E., P. H. Nielsen, H. J. Albrechtsen, and T. H. Christensen.** 1992. Importance of unattached bacteria and bacteria attached to sediment in determining potentials for degradation of xenobiotic organic contaminants in an aerobic aquifer. Appl. Environ. Microbiol. **58:**3020–3026.
19. **Holmes, A. J., A. Costello, M. E. Lidstrom, and J. C. Murrell.** 1995. Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. FEMS Microbiol. Lett. **132:**203–208.
20. **Horz, H. P., M. T. Yimga, and W. Liesack.** 2001. Detection of methanotroph diversity on roots of submerged rice plants by molecular retrieval of *pmoA*, *mmoX*, *mxaF*, and 16S rRNA and ribosomal DNA, including *pmoA*-based terminal restriction fragment length polymorphism profiling. Appl. Environ. Microbiol. **67:**4177–4185.
21. **Hosono, S., A. F. Faruqi, F. B. Dean, Y. Du, Z. Sun, X. Wu, J. Du, S. F. Kingsmore, M. Egholm, and R. S. Lasken.** 2003. Unbiased whole-genome amplification directly from clinical samples. Genome Res. **13:**954–964.

22. **Hyman, M. R., S. A. Russell, R. L. Ely, K. J. Williamson, and D. J. Arp.** 1995. Inhibition, inactivation, and recovery of ammonia-oxidizing activity in cometabolism of trichloroethylene by *Nitrosomonas europaea*. Appl. Environ. Microbiol. **61:**1480–1487.
23. **Idaho National Engineering and Environmental Laboratory.** 2001. Fiscal year 2000 groundwater monitoring annual report, Test Area North, Operable Unit 1-07B. Idaho National Engineering and Environmental Laboratory, Idaho Falls, Idaho.
24. **Koh, S.-C., J. P. Bowman, and G. S. Saylor.** 1993. Soluble methane monooxygenase production and trichloroethylene degradation by a type I methanotroph *Methylomonas methanica*. Appl. Environ. Microbiol. **59:**960–967.
25. **Lage, J. M., J. H. Leamon, T. Pejovic, S. Hamann, M. Lacey, D. Dillon, R. Segraves, B. Vossbrinck, A. Gonzalez, D. Pinkel, D. G. Albertson, J. Costa, and P. M. Lizardi.** 2003. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. Genome Res. **13:**294–307.
26. **Lasken, R. S., and M. Egholm.** 2003. Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. Trends Biotechnol. **21:**531–535.
27. **Lehman, R. M., F. S. Colwell, and G. A. Bala.** 2001. Attached and unattached microbial communities in a simulated basalt aquifer under fracture- and porous-flow conditions. Appl. Environ. Microbiol. **67:**2799–2809.
28. **Lehman, R. M., and S. P. O'Connell.** 2002. Comparison of extracellular enzyme activities and community composition of attached and free-living bacteria in porous medium columns. Appl. Environ. Microbiol. **68:**1569–1575.
29. **Lehman, R. M., S. P. O'Connell, A. Banta, J. K. Fredrickson, A. L. Reysenbach, T. L. Kieft, and F. S. Colwell.** 2004. Microbiological comparison of core and groundwater samples collected from a fractured basalt aquifer with that of dialysis chambers incubated *in situ*. Geomicrobiol. J. **21:**169–182.
30. **Lieberman, R. L., D. B. Shrestha, P. E. Doan, B. M. Hoffman, T. L. Stemmler, and A. C. Rosenzweig.** 2003. Bioinorganic chemistry special feature: purified particulate methane monooxygenase from *Methylococcus capsulatus* (Bath) is a dimer with both mononuclear copper and a copper-containing cluster. Proc. Natl. Acad. Sci. USA **100:**3820–3825.
31. **Little, C. K., A. V. Palumbo, S. E. Herbes, M. E. Lidstrom, R. L. Tyndall, and P. J. Gilmer.** 1988. Trichloroethylene biodegradation by a methane-oxidizing bacterium. Appl. Environ. Microbiol. **45:**951–956.
32. **Maltoni, C., G. Lefemine, C. Cotti, and G. Perino.** 1988. Long-term carcinogenicity bioassays on trichloroethylene administered by inhalation to Sprague Dawley rats and Swiss and B6C3F1 mice, p. 33–36. *In* C. Maltoni and I. J. Selikoff (ed.), Living in a chemical world. New York Academy of Science, New York.
33. **Marion, M. J., and S. Boivin-Angele.** 1999. Vinyl chloride-specific mutations in humans and animals. IARC Sci. Publ. **150:**315–324.
34. **McDonald, I. R., E. M. Kenna, and J. C. Murrell.** 1995. Detection of methanotrophic bacteria in environmental samples with the PCR. Appl. Environ. Microbiol. **61:**116–121.
35. **Melnick, R. L.** 2002. Carcinogenicity and mechanistic insights on the behavior of epoxides and epoxide-forming chemicals. Ann. N. Y. Acad. Sci. **982:**177–189.
36. **Miguez, C. B., D. Bourque, J. A. Sealy, C. W. Greer, and D. Groleau.** 1997. Detection and isolation of methanotrophic bacteria possessing soluble methane monooxygenase (sMMO) genes using the polymerase chain reaction (PCR). Microb. Ecol. **33:**21–31.
37. **Miller, R. E., and F. P. Guengerich.** 1983. Metabolism of trichloroethylene in isolated hepatocytes, microsomes and reconstituted enzyme systems containing cytochrome P-450. Cancer Res. **43:**1145–1152.
38. **Minin, V., Z. Abdo, P. Joyce, and J. Sullivan.** 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. **52:**674–683.
39. **Murrell, J. C., B. Gilbert, and I. R. McDonald.** 2000. Molecular biology and regulation of methane monooxygenase. Arch. Microbiol. **173:**325–332.
40. **Newby, D. T., D. W. Reed, L. M. Petzke, A. L. Igoe, M. E. Delwiche, F. F. Roberto, J. P. McKinley, M. J. Whiticar, and F. S. Colwell.** 2004. Diversity of methanotroph communities in a basalt aquifer. FEMS Microbiol. Ecol. **48:**333–344.
41. **Nguyen, H. H., S. J. Elliott, J. H. Yip, and S. I. Chan.** 1998. The particulate methane monooxygenase from *Methylococcus capsulatus* (Bath) is a novel copper-containing three-subunit enzyme. Isolation and characterization. J. Biol. Chem. **273:**7957–7966.
42. **Oldenhuis, R., R. L. Vink, D. B. Janssen, and B. Witholt.** 1989. Degradation of chlorinated aliphatic hydrocarbons by *Methylosinus trichosporium* OB3b expressing soluble methane monooxygenase. Appl. Environ. Microbiol. **55:**2819–2826.
43. **Page, R. D.** 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12:**357–358.
44. **Semrau, J. D., D. Zolandz, M. E. Lidstrom, and S. I. Chan.** 1995. The role of copper in the *pMMO* of *Methylococcus capsulatus* bath: a structural vs. catalytic function. J. Inorg. Biochem. **58:**235–244.
45. **Shigematsu, T., S. Hanada, M. Eguchi, Y. Kamagata, T. Kanagawa, and R. Kurane.** 1999. Soluble methane monooxygenase gene clusters from trichloroethylene-degrading *Methylomonas* sp. strains and detection of methanotrophs during in situ bioremediation. Appl. Environ. Microbiol. **65:**5198–5206.
46. **Song, D. L., M. E. Conrad, K. S. Sorenson, and L. Alvarez-Cohen.** 2002. Stable carbon isotope fractionation during enhanced in situ bioremediation of trichloroethene. Environ. Sci. Technol. **36:**2262–2268.
47. **Sorenson, K. S., L. N. Peterson, R. E. Hinchee, and R. L. Ely.** 2000. An evaluation of aerobic trichloroethene attenuation using first-order rate estimation. Bioremediat. J. **4:**337–357.
48. **Storck, W.** 1987. Chlorinated solvent use hurt by federal rules. Chem. Eng. News **65:**11.
49. **Tsien, H. C., G. A. Brusseau, R. S. Hanson, and L. P. Waclett.** 1989. Biodegradation of trichloroethylene by *Methylosinus trichosporium* OB3b. Appl. Environ. Microbiol. **55:**3155–3161.
50. **Tsien, H. C., and R. S. Hanson.** 1992. Soluble methane monooxygenase component B gene probe for identification of methanotrophs that rapidly degrade trichloroethylene. Appl. Environ. Microbiol. **58:**953–960.
51. **Webb, J. S., M. Givskov, and S. Kjelleberg.** 2003. Bacterial biofilms: prokaryotic adventures in multicellularity. Curr. Opin. Microbiol. **6:**578–585.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Aug. 2003, p. 4927–4934 Vol. 69, No. 8
0099-2240/03/$08.00+0 DOI: 10.1128/AEM.69.8.4927–4934.2003

Metagenomic Profiling: Microarray Analysis of an Environmental Genomic Library

Jonathan L. Sebat,[1]† Frederick S. Colwell,[2] and Ronald L. Crawford[1]*

Environmental Research Institute, University of Idaho, Moscow, Idaho 83844-1052,[1] and Idaho National Engineering and Environmental Laboratory, Idaho Falls, Idaho 83415[2]

Received 3 March 2003/Accepted 30 May 2003

Genomic libraries derived from environmental DNA (metagenomic libraries) are useful for characterizing uncultured microorganisms. However, conventional library-screening techniques permit characterization of relatively few environmental clones. Here we describe a novel approach for characterization of a metagenomic library by hybridizing the library with DNA from a set of groundwater isolates, reference strains, and communities. A cosmid library derived from a microcosm of groundwater microorganisms was used to construct a microarray (COSMO) containing ~1-kb PCR products amplified from the inserts of 672 cosmids plus a set of 16S ribosomal DNA controls. COSMO was hybridized with Cy5-labeled genomic DNA from each bacterial strain, and the results were compared with the results for a common Cy3-labeled reference DNA sample consisting of a composite of genomic DNA from multiple species. The accuracy of the results was confirmed by the preferential hybridization of each strain to its corresponding rDNA probe. Cosmid clones were identified that hybridized specifically to each of 10 microcosm isolates, and other clones produced positive results with multiple related species, which is indicative of conserved genes. Many clones did not hybridize to any microcosm isolate; however, some of these clones hybridized to community genomic DNA, suggesting that they were derived from microbes that we failed to isolate in pure culture. Based on identification of genes by end sequencing of 17 such clones, DNA could be assigned to functions that have potential ecological importance, including hydrogen oxidation, nitrate reduction, and transposition. Metagenomic profiling offers an effective approach for rapidly characterizing many clones and identifying the clones corresponding to unidentified species of microorganisms.

Microorganisms contribute significantly to the earth's biological diversity, yet relatively few of the microorganisms present in nature have been cultured and characterized. It is generally accepted that less than 1% of bacteria and fungi present in most habitats have been cultivated for study in pure culture (2). Although direct analysis of environmental DNA samples by PCR is effective for showing the presence of uncultured microorganisms, biases in primer specificity and amplification of different targets prevent full recognition of microbial diversity (31, 37, 40, 41). Thus, new approaches to examination of community genomes are needed.

The use of large-insert genomic libraries is a powerful approach for isolating DNA sequences from complex mixtures of uncultured microorganisms. Direct cloning of DNA from environmental samples makes it possible to avoid some of the biases of cultivation and PCR. In addition, genomic fragments that are >100 kb long can be obtained, and they provide significant functional and taxonomic information about the organisms from which they were derived. Such metagenomic libraries have been used to identify novel genes from uncultivated species of archaea, bacteria, and viruses that are responsible for significant ecosystem processes (4, 5, 8, 12) and to isolate enzymes that are involved in biosynthesis of novel pharmaceuticals (7, 15, 24, 41, 42) or have other industrial uses (11, 16, 17, 26, 33).

* Corresponding author. Mailing address: Environmental Research Institute, University of Idaho, Moscow, ID 83844-1052. Phone: (208) 885-6580. Fax: (208) 885-5741. E-mail: crawford@uidaho.edu.

† Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.

Given the immense uncultivated and uncharacterized metabolic diversity in the environment, one would need to sequence relatively few bacterial artificial chromosome (BAC) or cosmid clones to discover fundamentally interesting sets of genes. If modern genomic techniques can be used to carry out more comprehensive surveys of metagenomic libraries, our understanding of natural genetic diversity should be greatly enhanced.

Screening a genomic library can be done a number of different ways. Typically, screening involves colony hybridization with a probe of interest, which yields information one gene at a time. Bioassays have been developed to screen libraries for genes involved in the production of specific enzymes (11, 16, 17, 26, 33) or natural products (15, 24, 42); however, this approach relies on the fortuitous expression of heterologous DNA by the library host strain. High-throughput end sequencing of BAC clones has been used to accelerate various single-genome projects (39), and it is currently being used to characterize some environmental DNA libraries (8).

Although the speed and effectiveness of brute-force sequencing are constantly improving, it is not yet practical to assemble a complete bacterial genome from a metagenome. There is still a need for new functional genomic approaches that systematically yield information about many of the elements in a metagenomic library. These new approaches should ideally allow us to identify the organism from which each clone came, to determine some functional characteristics of various clones, and to identify many more novel uncultivated bacteria.

We sought to develop a practical approach that would provide a large amount of information about the microbial com-

munity from a limited set of clones. The purpose of our approach was to classify many of our cosmids and to identify a few candidates for sequencing rather than to undertake a major sequencing and assembly project. Our method involves hybridization of the library with genomic DNA of various reference strains and bacterial isolates from the community under study. In addition, DNA derived from as-yet-uncultivated organisms can be identified by hybridization with metagenomic DNA.

Metagenomic profiling. Metagenomic profiling is classification of clones based on hybridization of insert DNA to the genomes of bacterial isolates, reference strains, and environmental DNA.

DNA microarray technology has become an important tool for determining the gene contents of entire genomes and measuring the expression of genes (18). High-density arrays are effective for quantitative detection of genes in complex samples. Thus, microarrays are a promising technique for characterization of genes in environments such as soil and water (3, 9, 34, 44, 45). However, the use of microarrays has been limited to 16S rRNA markers or a relatively small set of functional genes, and no practical approach has been developed to specifically target the unculturable majority of the species in the environment.

We used a microarray platform to screen a metagenomic library with whole microbial genomes and community genomes (Fig. 1). The microarray (COSMO) contained ~1-kb PCR products amplified from the inserts of 672 cosmids along with a set of controls (16S ribosomal DNA [rDNA] probes). From an environmental sample (the same sample from which the library was derived), numerous bacterial isolates were obtained. Genomic DNA was purified from each environmental isolate. In addition, metagenomic DNA was purified directly from the mixed population, which was done without cultivation of bacteria. Each test genome was labeled with Cy5-dCTP and probed with COSMO. In order to subtract any signal that may have come from nonspecific hybridization, in each experiment we used two-color hybridization, in which each test genome was compared to a reference sample of common bacterial DNA. The reference DNA consisted of a pooled sample of genomic DNA from 14 species of bacteria (which effectively diluted the strain-specific genes and enriched common sequences). The reference DNA was given a different label (Cy3), and equal amounts of test and reference DNA were combined and hybridized to COSMO. We refer to this approach as comparative genomic hybridization (CGH). CGH was repeated for all environmental isolates, as well as for the metagenomic DNA sample(s). Positive results were determined based on a Cy5/Cy3 ratio greater than 1 (>0 on a $\log_2$ scale). As a result, we obtained a profile for each clone in the metagenomic library (i.e., a graphical representation of its hybridization to one or more species of bacteria). Clones that were specific to a test strain or community hybridized only to those DNA samples. Clones that contained a conserved sequence within their corresponding PCR amplicons hybridized to the genomes of multiple species.

MATERIALS AND METHODS

Bacterial strains, media, and culture conditions. To evaluate our approach, we needed a microbial community that could be manipulated in the laboratory.

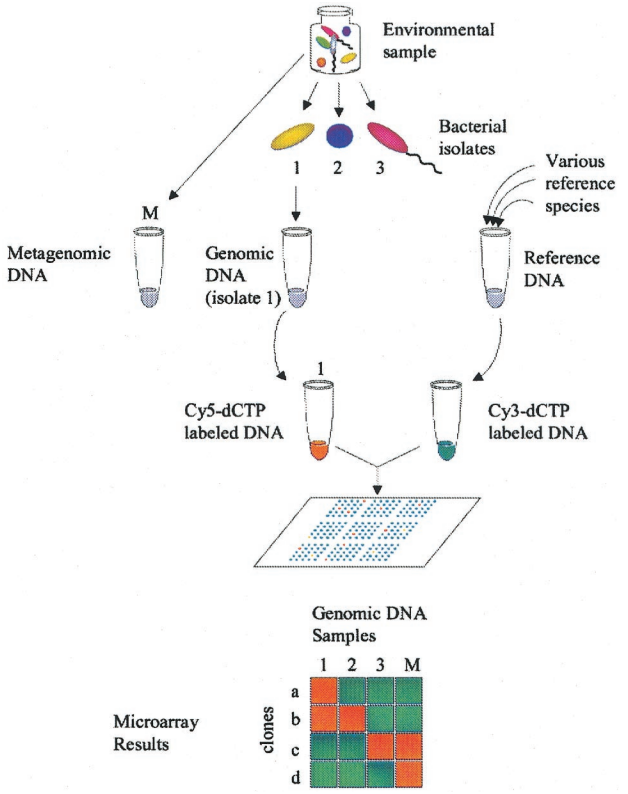


FIG. 1. Metagenomic profiling. Genomic DNA is purified from various bacterial isolates, as well as from a mixed population. A reference sample of common bacterial DNA is created by pooling the genomic DNA of many strains. In each CGH experiment, the genome of each strain or community is analyzed in comparison to the reference DNA. Some examples of informative results are shown. A clone of environmental DNA may correspond to only bacterial strain 1 (clone a), multiple strains (clone b), the metagenome and a bacterial isolate (clone c), or only the metagenome (clone d).

A stable community was developed from an inoculum of biofilm material collected from the Snake River Plain Aquifer in southeastern Idaho. Biofilm was collected from the basalt aquifer by suspending a basket containing 100 ml of ceramic beads in an open borehole at a depth of 73 m adjacent to a zone of high groundwater flow. After 80 days, the basket was retrieved, and the beads were immersed in 100 ml of sterile phosphate-buffered saline. Cells were collected by gently vortexing the submerged beads for 15 min. One milliliter of a cell suspension was used to inoculate triplicate flasks containing 100 ml of minimal succinate medium and 100 ml of glass beads. Minimal succinate medium contained (per liter of deionized water) 6.0 g of K_2HPO_4, 3.0 g of KH_2PO_4, 1.0 g of $(NH_4)_2SO_4$, and 4.0 g of succinic acid. The pH of the medium was adjusted to 7.5 with 10 M NaOH. One milliliter of sterile 1 M $MgSO_4$ and 1.0 ml of sterile 45 mM $CaCl_2$ were added after autoclaving. Microcosms were developed by incubating the flasks without shaking for 1 week at 30°C. After incubation, the cells were suspended by gently vortexing the flask and decanting the medium into two 50-ml polypropylene tubes, and cells were collected by centrifugation and resuspended in 1.0 ml of phosphate-buffered saline. Approximately 10^{10} cells were used for construction of a cosmid library, and the remaining cells were stored in glycerol at −80°C.

To obtain community DNA that were to be used for microarray target samples, a second enrichment of the microcosm was carried out under the same conditions by using a 0.5% inoculum of frozen stock. After incubation, two different fractions of cells were collected (pellicle and planktonic). First, the pellicle was removed with a sterile spatula and placed in a 50-ml polypropylene tube, and then the remaining cells were suspended by gently vortexing the flask, decanting the medium into two 50-ml polypropylene tubes, and collecting the cells by centrifugation. DNA from the samples described above was probed with COSMO in order to identify clones corresponding to microbial strains that were

enriched in the mixed culture but that we failed to isolate in pure culture. We used the unisolated fraction of the microcosm to represent the uncultured microorganisms in the environment.

From the original microcosm and all subsequent enrichments, bacterial species were isolated by plating serial dilutions of liquid cultures onto plates containing tryptic soy agar. Bacterial strains were identified by selecting colonies with unique morphology that appeared during a 7-day incubation at 30°C. Ten different strains were obtained (see Fig. 4) along with the *Pseudomonas aeruginosa* and *Staphylococcus aureus* reference strains. To determine the identity of each isolate, the 16S rRNA gene was amplified by PCR from a genomic DNA template by using eubacterial primers 27F (20) and 907R (29). 16S amplicons were sequenced by using the 27F primer.

Library construction. A cosmid library was constructed from the genomic DNA of the original mixed bacterial enrichment. Bacterial cells were embedded in agarose, and genomic DNA was purified by agarose-embedded cell lysis, followed by partial digestion of the agarose plugs with *Sau*3AI as described by Stein et al. (35). The metagenomic library was constructed by using a SuperCos I cosmid kit (Stratagene) according to the manufacturer's protocols. Clones were picked randomly into 96-well plates containing Luria-Bertani medium (44) supplemented with 100 mg of ampicillin per liter and 0.1 volume of 10× Hogness buffer (40% glycerol, 36 mM K_2HPO_4, 13 mM KH_2PO_4, 20 mM trisodium citrate, 10 mM $MgSO_4$ in deionized water). After overnight incubation, the library was stored at −80°C. Plasmids were purified by using a REAL prep 96 kit (Qiagen) and a BioRobot 3000 (Qiagen) liquid-handling system. When we examined *Xho*I restriction digests of 10 clones by agarose gel electrophoresis, we observed no duplicate clones, and we determined that the average insert size of the cosmids was ~40 kb. The results presented below verify that COSMO represented much of the microcosm's diversity. We did not attempt to characterize the species represented in the cosmid library prior to fabrication of COSMO. The task of amplifying 16S rRNA genes from a pool of cosmids was confounded by the presence of contaminating *Escherichia coli* genomic DNA in the plasmid preparation. (While the manuscript was being reviewed, Liles et al. [21] published a new procedure for eliminating *E. coli* 16S rDNA from pools of purified BACs.)

Array fabrication. All microarray experiments were performed with COSMO, a DNA microarray containing end fragments of 672 cosmids selected randomly from the metagenomic library plus a set of reference genes (16S rDNA markers from several microcosm isolates [Fig. 2]). Cosmid end fragments were produced by a thermal asymmetric interlaced PCR method (22). Briefly, the thermal asymmetric interlaced PCR method involved sequential cycles of linear amplification of insert DNA from the T7 end of the vector with nested primers, followed by exponential amplification of the specific product by random priming, which resulted in PCR products that were 200 to 2,000 bp long. High-throughput PCR was performed with an MBS 384S thermocycler system (ThermoHybaid). 16S rRNA gene controls were also amplified by PCR. The quality and quantity of DNA were confirmed by agarose gel electrophoresis. PCR products were purified by using 384-well filter plates (Millipore) and were resuspended in 15 μl of 1× Spotting Solution Plus (Telechem) to obtain a final DNA concentration of 100 to 200 ng/μl. Each sample was spotted in duplicate on SuperAmine slides (Telechem) by using a Microgrid arrayer (BioRobotics). Slide cross-linking, washing, and blocking steps were carried out by using the manufacturer's protocols (http://arrayit.com/PDF/Super_Microarray_Substrates.pdf).

DNA preparation, labeling, and hybridization. Genomic target DNA was purified from bacterial isolates and mixed cultures as described by Wilson (43). The reference sample was prepared by mixing equal amounts of genomic DNA from the 10 species of bacterial isolates that were used in this study and from the reference organisms *E. coli*, *P. aeruginosa*, *S. aureus*, and *Bacillus subtilis*. Fluorescently labeled target DNA was made as described by Pollack et al. (30). Briefly, 2 μg of target DNA was digested completely with *Msp*I and purified by ethanol precipitation. Prior to labeling, 10 ng of *salT* (14) (a rice gene used as an internal standard) was added to the sample. Target DNA was labeled by incorporation of Cy5-dCTP (for test DNA) or Cy3-dCTP (for reference DNA) (Amersham Pharmacia) by random primer synthesis (BioPrime labeling kit; Invitrogen). Labeled target DNA was purified with CHROMA SPIN+ TE-30 gel filtration columns (Clontech). Test DNA and reference DNA were combined with 40 μg of human Cot-1 DNA and 100 μg of salmon sperm genomic DNA and reduced to a volume of 5 μl by using Microcon YM-30 concentrators (Millipore); 20 μl of 1.25× unihyb hybridization buffer (Telechem) was added to the target DNA mixture, and the target was preannealed to blocking DNA by boiling the preparation for 1.5 min, followed by 30 min of incubation at 37°C. Twenty-five microliters of probe was used per slide. Hybridization was performed for 8 h at 65°C. Posthybridization washing was performed by using the slide manufacturer's protocol.

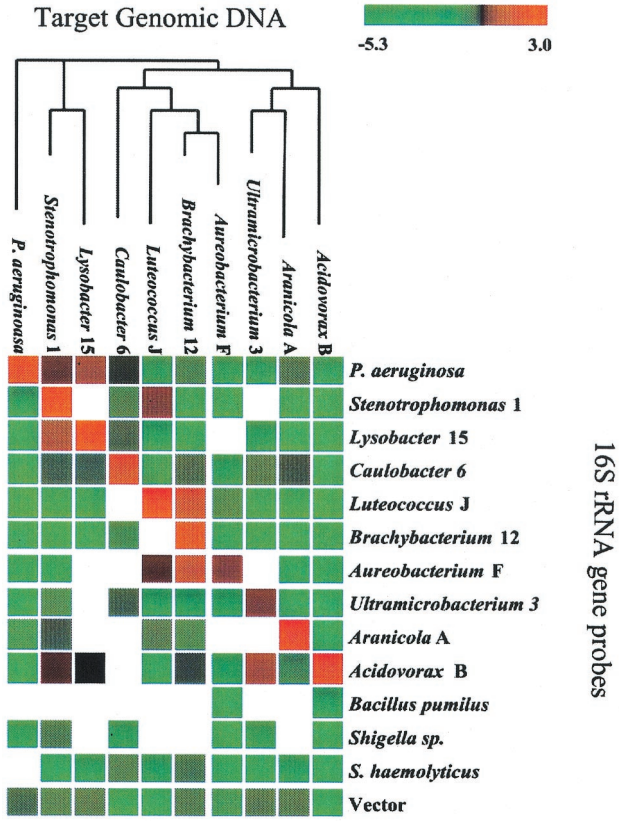


FIG. 2. Microarray validation with 16S rDNA controls. The data are the results obtained with rRNA probes in 10 different experiments (one replicate per strain). The results of each experiment are presented as a column of color-coded values. The colors, ranging from green to red, correspond to the $\log_2$ of the Cy5/Cy3 ratio (see key at the top). Preferential hybridization of the test strain to a rRNA gene probe is indicated by a positive $\log_2$ ratio (red). Results with a signal-to-noise ratio of less than 3 are indicated by blank squares. Species are sorted horizontally and vertically according to their phylogenetic relationships (as determined by Clustal W alignment of rRNA sequences) in order to show the amount of cross-hybridization that occurs between the sequences of related strains. The last four probes shown are additional negative controls.

Data analysis. Arrays were scanned with an Axon 4000 scanner, and fluorescence measurements were obtained by using Genepix Pro 3.0 software (Axon). Data sets were filtered for spots with a signal-to-noise ratio greater than 3.0 by using Microsoft Excel. Results are reported below as $\log_2$ of the Cy5/Cy3 ratio. Data analysis and graphical display were done with Expression NTI (Informax Inc.) in the following manner: (i) all clones that failed to produce a positive result ($\log_2$, >0) in duplicate experiments with at least one target genome were removed from the data set, (ii) clones were clustered by complete linkage of genes by using the correlation coefficient, and (iii) experiments were sorted according to the phylogenetic relationships of test strains as determined by Clustal W alignment of 16S rRNA sequences.

Analysis of cosmid end sequences. The inserts of select clones were sequenced from both ends of the multiple cloning site by using the T3 and T7 primers. The plasmid template was purified from 500-ml cultures of *E. coli* by using a large-construct kit (Qiagen). A sequencing reaction mixture (total volume, 20 μl) was prepared by combining 1 μg of plasmid DNA with 0.32 μl of a 10 μM primer T3 or T7 stock solution in Tris-EDTA buffer and 8.0 μl of a Big Dye mixture. PCR was performed with a PTC-100 thermocycler (MJ Research) for 80 cycles (94°C for 30 s, 47°C for 15 s, and 60°C for 4 min). Reaction mixtures were purified with DTR gel filtration columns (Edge Biosystems). Nucleotide sequences (average size, 450 bp) were determined with an ABI 3100 DNA sequencer.

The potential functions and phylogenetic affinities of cosmid end sequences were determined by performing a nucleotide and translated-protein search of

TABLE 1. Potential genes observed in cosmid end sequences[a]

Clone	Gene in T7 end	Gene in T3 end[b]
2H6	Transposase (*Burkholderia fungorum*)	HP (*Chloroflexus aurantiacus*)
3C2	ABC transporter (*Pseudomonas aeruginosa*)	HP (*Burkholderia fungorum*)
3F7	NS	ABC transporter (*Ralstonia metallidurans*)
3G10	Replication initiator protein *dnaA* (*Ralstonia metallidurans*)	HP (*Sinlorhizobium meliloti*)
3H9	Unknown (*Azotobacter vinelandii*)	HP (*Desulfitobacterium* sp.)
5D6	Denitrifying NorE and NorF genes (*Pseudomonas stutzeri*)	NS
5D8	50S ribosomal subunit protein L32 (*Ralstonia solanacearum*)	HP (*Desulfovibrio desulfuricans*)
5D11	Short-chain dehydrogenase (*Rhizobium* sp. strain NGR234)	HP (*Halobacterium* sp. strain NR)
5G2	Putative [NiFe] hydrogenase (*Streptomyces averimitilis*)	NS
6A7	Transposase and tRNA synthase (*Ralstonia metallidurans*)	Probable serine hydroxymethyl transferase (*Microbulbifer degradans*)
6A12	Insertion sequence IS*1051-X* (*Xanthomonas oryzae*)	NS
6D10	Transposase (*Ralstonia metallidurans*)	Cytochrome *c*-like protein (*Burkholderia fungorum*)
6F10	Probable organic solvent resistance protein (*Ralstonia solanacearum*)	Sensory transduction histidine kinase (*Magnetospirillum magnetotacticum*)
7A12	Integrase-like protein (*Xanthomonas axonopodis*)	Peptidyl-prolyl *cis-trans* isomerase (*Ralstonia solanacearum*)
7B11	Cation-transporting ATPase (*Mycobacterium tuberculosis*)	HP (*Geobacter metallireducens*)
7H7	HP (*Novosphingobium aromaticivorans*)	HP (*Halobacterium* sp. strain NRC-1)

[a] Sequences were read from the T7 and T3 ends of each clone. Each result is a consensus of all significant results from nucleotide and translated protein searches of the GenBank database. The species or strain corresponding to the most significant result at the nucleotide level is indicated in parentheses.
[b] HP, hypothetical protein; NS, no significant homology.

GenBank by using the Basic Local Alignment Search Tool (BLAST) (1). The potential function of a given sequence was determined by examination of all homologous sequences with expect values of <1e-2. A particular function was assigned if (i) a consensus was apparent among the best hits and (ii) there was no disagreement between the consensus of the nucleotide results and the translated-protein results. We also listed the species corresponding to the nucleotide best hit; when no significant nucleotide matches were observed, we listed the species corresponding to the protein best hit (Table 1).

RESULTS

Data validation. A single self-versus-self hybridization was performed by hybridizing COSMO with two samples of *Acidovorax* B genomic DNA that had been prepared and labeled separately, one with Cy5 and the other with Cy3. Seventy-six probes corresponding to 43 different cosmids passed our filtering criteria (see Materials and Methods). The values for the Cy5/Cy3 ratio ranged from 0.77 to 1.19 (mean, 1.00; standard deviation, 0.12) (raw data not shown). Based on this experiment, the performance of different probes and dyes should have caused less than a 1.3-fold deviation from the mean.

The sensitivity and specificity of our microarray analysis were evident from the results obtained for controls and reference genes (Fig. 2) in our CGH experiments with individual genomes (see below). 16S rRNA genes from several microcosm isolates were obtained by PCR and included in COSMO as controls. Each rRNA gene served as a positive control for its corresponding species and as a negative control for distantly related species. Additional negative controls are shown last in the figures. The results indicated that our hybridization conditions allowed identification of strain-specific genes. In all but one case, genomic DNA of each strain hybridized preferentially to its corresponding rRNA probe (yielding the highest ratio). In most cases, some hybridization to a related strain was observed, but the level of hybridization was lower. In the case of *Ultramicrobacterium* 3, genomic DNA hybridized nearly equally to the *Ultramicrobacterium* 3 and *Acidovorax* B rRNA probes. No positive results were obtained with the *Bacillus* sp., *Shigella* sp., and *Staphylococcus* sp. negative controls or with the cosmid vector.

Hybridization with individual genomes. Various reference strains and microcosm isolates were used individually as target DNA in experiments to locate clones related to these organisms. The clustered microarray results for various microcosm isolates and reference strains revealed distinct classes of DNA that corresponded to individual strains or groups of bacteria (Fig. 3). Numerous strain-specific clones were apparent. The data also revealed examples in which clones hybridized to multiple related organisms (indicative of conserved genes). Based on these patterns, we classified some clones as members a particular species, genus, or branch (Fig. 3).

At the bottom of Fig. 3, some results appear to be scrambled (i.e., distinct patterns are not easily distinguished). The profiles observed for this group of clones are not consistent with profiles of strain-specific and conserved clones. In general, these clones appeared to cross-hybridize between species to a greater extent. A variety of different patterns were observed, and for the sake of simplicity they were not labeled.

Classifying uncultured DNA. In the experiment described above (Fig. 3), 156 probes (clones) passed our filtering process. The remaining 524 clones failed to produce a positive $\log_2$ ratio with any test strain or failed to produce any significant signal. Some of these clones may have corresponded to organisms present in the microcosm that we failed to isolate in pure culture. We considered the unisolated organisms in our microcosm to be analogous to uncultured microbes in the environment. To identify cosmids derived from such organisms, we performed a similar experiment using genomic DNA extracted directly from the mixed bacterial population as the test DNA (the same reference DNA was used). The results of two metagenomic CGH experiments were added to the data set prior to clustering (Fig. 4). The results identified a number of clones that were present in the community and not in our catalog of isolates (Fig. 4A). Such clones were classified as uncultured.

A single experiment, such as the one described above, yielded a spectrum of ratios. It is likely that the clones that yielded a $\log_2$ ratio of 1 corresponded to a different organism than the clones that yielded a $\log_2$ ratio of 6. The uncultured class could be separated into subgroups if there were clear differences in the abundance of different genes in the community, but it was not obvious where to draw the line between one organism and another.

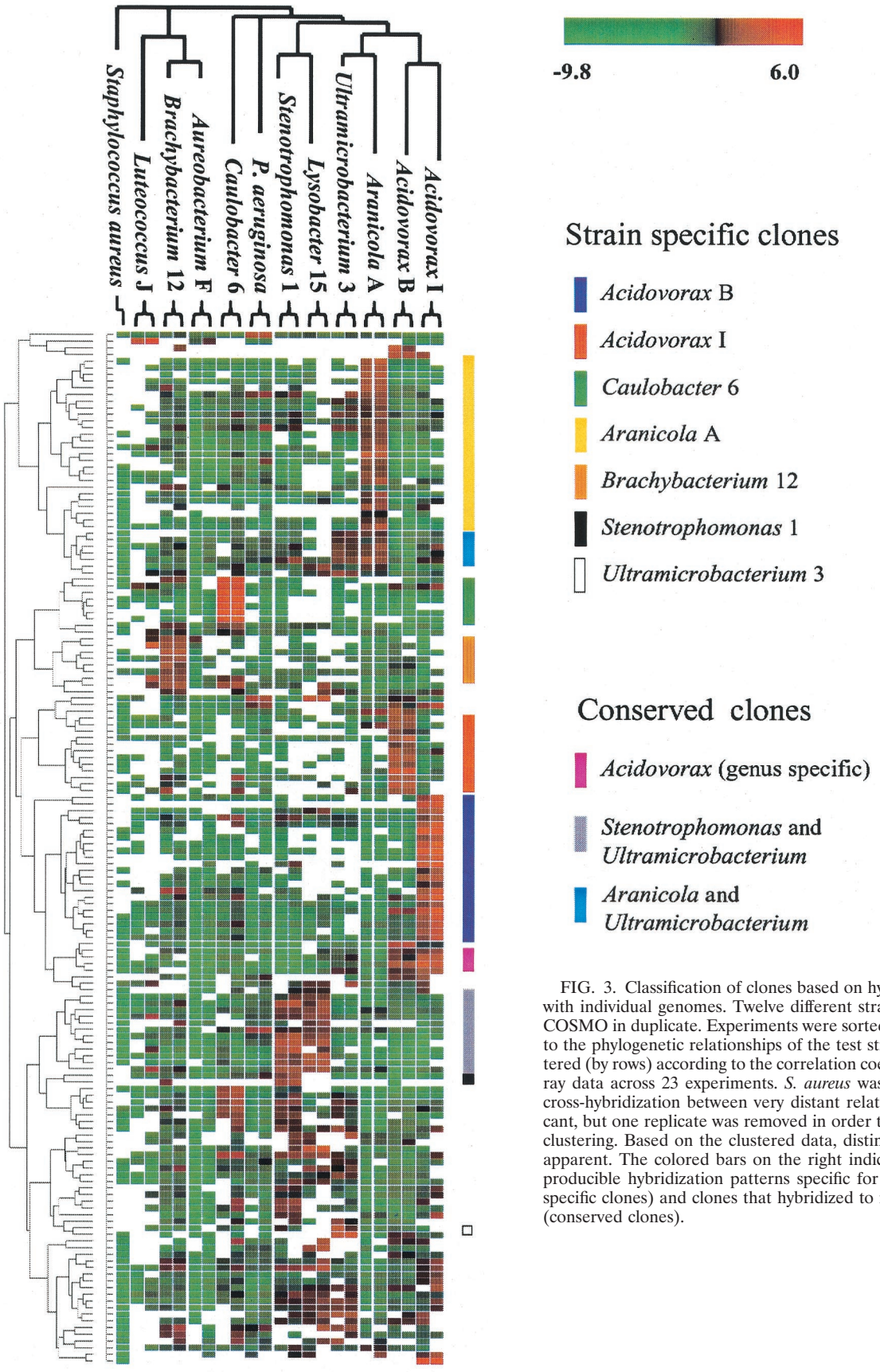


FIG. 3. Classification of clones based on hybridization of COSMO with individual genomes. Twelve different strains were analyzed with COSMO in duplicate. Experiments were sorted horizontally according to the phylogenetic relationships of the test strains. Clones were clustered (by rows) according to the correlation coefficients of the microarray data across 23 experiments. *S. aureus* was included to show that cross-hybridization between very distant relatives may not be significant, but one replicate was removed in order to minimize its effect on clustering. Based on the clustered data, distinct classes of clones are apparent. The colored bars on the right indicate the clones with reproducible hybridization patterns specific for one test strain (strain-specific clones) and clones that hybridized to multiple related species (conserved clones).

In an attempt to resolve the uncultivated community in slightly better detail, the microcosm cells were fractionated into two types, pellicle cells and free-swimming cells, which were analyzed separately. All uncultured clones (Fig. 4A) were found to have the same even distribution between planktonic and pellicle cells; therefore, we were not able to determine if these clones corresponded to multiple species. However, we observed that clones corresponding to some of the cultivable species had distinct distribution patterns. *Acidivorax* I was apparently distributed throughout the microcosm (Fig. 4B), *Acidivorax* B was present primarily in the pellicle (Fig. 4C), and *Caulobacter* 6 was not abundant in either fraction (Fig. 4D). These clusters were labeled and presented as examples of clones from different species that, in principle, could be distinguished based solely on community analysis.

Sequence analysis of uncultivated DNA. In order to identify functional characteristics of uncultured microorganisms from the microcosm, insert DNA from each of 17 random clones from cluster A (Fig. 4A) was sequenced from the T7 and T3 primer sites that flanked the insert. A nucleotide and translated-protein search of GenBank was performed with each cosmid end sequence (Table 1). Among 34 sequences, 18 functional genes were identified, 10 hits were obtained with genes of unknown function, and 6 sequences yielded no significant result. The great majority of the sequences were found to be significantly similar to genes from members of the *Proteobacteria*, including seven genes from *Ralstonia* sp. Some of the sequences could be assigned to functions having ecological importance, including a putative [NiFe] hydrogenase, nitrate reduction, and several transposases. Four different insertion sequence elements (ISs) were observed in five clones, 2H6, 6A12, 6A7, 6D10, and 7A12 (sequences from 6A7 and 6D10 were from different positions of the same gene). All ISs that we identified occurred precisely in the T7 end fragment.

DISCUSSION

The experiments described above illustrate a useful technique for rapidly classifying DNA from a metagenomic library and, in the process, identifying conserved and divergent genome fragments from a particular community, genus, strain, or species and DNA corresponding to strains that have not been isolated in pure culture.

Hybridizing COSMO with single genomes and cluster analysis of the microarray results were effective for visualizing groups of clones that have unique patterns of hybridization to different bacterial genomes (Fig. 3). We identified clones that hybridized to a single strain, as well as clones that hybridized to related species. In this manner, we classified clones as strain specific or specific to broader phylogenetic group.

Many of the clones in the scrambled section of Fig. 3 have much broader specificity, and some cross-hybridize only between two distantly related strains. The latter finding may be of interest to researchers who wish to identify genes that have been transferred horizontally between species. However, before such claims are made based on hybridization between more divergent species, the correlation between homology and signal intensity must be examined more carefully.

Identification of clones corresponding to cultivable strains is also useful for eliminating clones that do not need to be examined in a culture-independent manner. This may in fact be the primary objective, in which case it would not be necessary to probe genomes one at a time, as we did. The cultivable genomes could be combined into pools in order to accelerate the process.

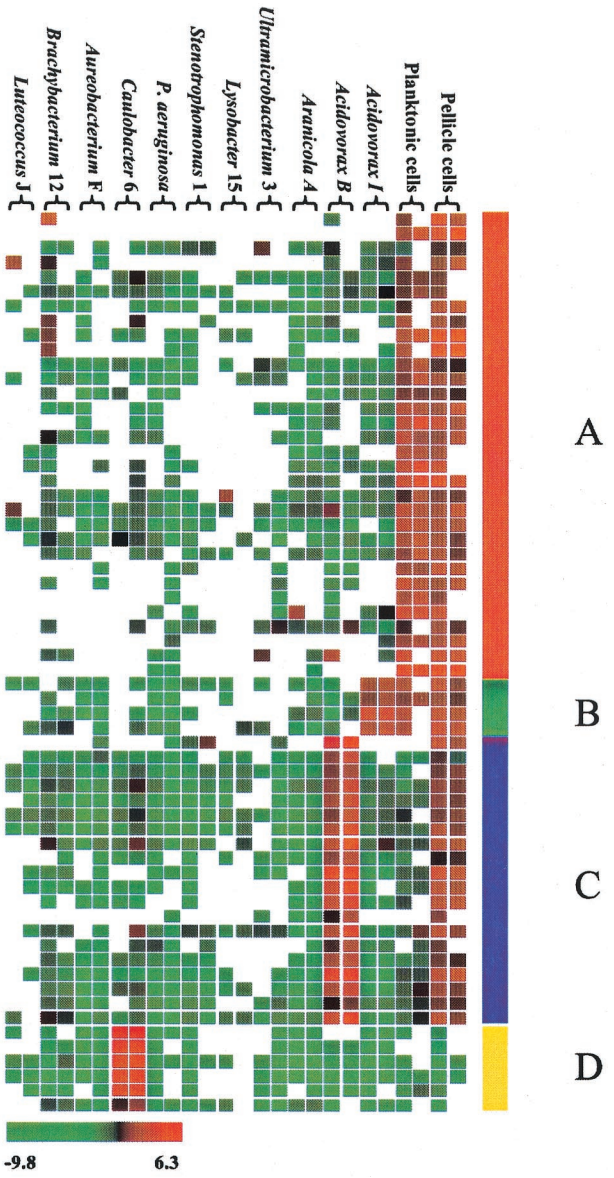


FIG. 4. Identification of clones corresponding to uncultivated organisms. CGH experiments were performed by using metagenomic DNA from different cell fractions, including free-swimming (planktonic) cells and cells accumulating at the surface (pellicle cells). When duplicate experiments were added to the original data set, it was possible to identify uncultivated DNA (A). In addition, probes classified by single-genome CGH could be used to track the distribution of an organism in the mixed population (B, C, and D).

The eventual identification of clones corresponding to uncultivated microorganisms is the most valuable aspect of this approach. We identified clones that hybridized uniquely to total community DNA. These clones corresponded to one or more species of bacteria that were enriched in mixed cultures but that we were unable to obtain by isolation on tryptic soy agar. End sequencing of 17 cosmids resulted in identification

of numerous sequences that had nucleotide similarity to the genes of *Ralstonia* spp. In addition, we identified a putative hydrogenase and genes involved in reduction of nitrate. Some species of bacteria, including *Ralstonia* spp. (6, 19, 32), have the ability to couple hydrogen oxidation to nitrate reduction. The putative hydrogenase gene and the *norE* homologue were not found in the same clone, and we could not confirm that they are linked any way. Such proof would require isolating both sets of genes on the same fragment of DNA. However, knowledge of the metabolic genes present in the uncultured population provides information that should be important for selectively culturing such organisms if they are indeed present.

Among the clones that hybridized only to community DNA, a remarkably high frequency of ISs was observed. We do not believe that this is a unique characteristic of the uncultivated species. ISs occur in a wide range of genera (25). A variety of ISs can occur in a single genome, and a single type of IS may have a copy number of >20 (38). We believe that the five occurrences of four different IS-like elements precisely in a clone's T7 end fragment (the probe actually printed on the array) may have reflected a greater sensitivity of our method for high-copy-number sequences.

Of course, in a mixed population, our method detects the most abundant genes. A common limitation of microarray studies is sensitivity. The detection limit for microarray analysis of soil samples is on the order of ~50 ng of a single bacterial genome (10); therefore, in a typical experiment one would detect only organisms that constitute at least 1/40 of the population. However, new techniques for uniform amplification of genomic DNA (27) and enrichment of unique sequences (23, 28) may be applied to environmental samples in order to access the single-copy genes of less abundant species.

We hope to further develop applications of the COSMO microarray for environmental samples. We demonstrate here that in addition to simple identification of an uncultured subset of clones, additional classes of organisms can be defined by comparing different environmental samples from the same habitat (Fig. 4B to D). Clones that have a unique distribution in the two fractions may constitute a separate phylogenetic class. By comparing the metagenomic profiles of a field site to a map of metabolic activities (13) or microbial species (36), clones whose distribution correlates with a biological process may be identified. In this manner, ecologically important genes for which there is not a specific probe or assay may be identified.

Systematic organization of a metagenomic library is essential for performing a more comprehensive study of a microbial community. We must continue to utilize modern techniques of genome analysis and adapt them to the study of complex mixed genomes, keeping in mind that the purpose of a genome-wide study is to accelerate the discovery of genes that are important for specific processes.

A study of the microbial metagenome need not seek truly comprehensive knowledge concerning all microbial genomes if it is possible to first negotiate the genomic landscape of a given site and find what is relevant and interesting. A practical approach to metagenomics is (i) to quickly identify familiar genes, (ii) to identify the unknowns, and (iii) to attempt to classify them based on knowledge such as where certain genes are present and what apparent linkages there are between different genes. Once a set of clones that is linked to a biological process is identified, the specific genes involved may be identified from the complete DNA sequences of the clones. Metagenomic profiling is an appropriate technique for these tasks. We believe that microarray studies in combination with DNA sequence analysis will be important tools for enhancing our understanding of earth's microbial diversity.

ACKNOWLEDGMENTS

We thank Cornelia Sawatzky for assistance with the preparation of the manuscript and Alan Caplan for supplying control DNA for the microarray.

This work was funded by a predoctoral fellowship from the Inland Northwest Research Alliance. Additionally, the project was supported by NIH grant P20 RR16454 from the BRIN Program of the National Center for Research Resources.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. **59:**143–169.
3. **Bavykin, S. G., J. P. Akowski, V. M. Zakhariev, V. E. Barsky, A. N. Perov, and A. D. Mirzabekov.** 2001. Portable system for microbial sample preparation and oligonucleotide microarray analysis. Appl. Environ. Microbiol. **67:** 922–928.
4. **Beja, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, and E. F. DeLong.** 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science **289:**1902–1906.
5. **Beja, O., E. N. Spudich, J. L. Spudich, M. Leclerc, and E. F. DeLong.** 2001. Proteorhodopsin phototrophy in the ocean. Nature **411:**786–789.
6. **Bernhard, M., E. Schwartz, J. Rietdorf, and B. Friedrich.** 1996. The *Alcaligenes eutrophus* membrane-bound hydrogenase gene locus encodes functions involved in maturation and electron transport coupling. J. Bacteriol. **178:** 4522–4529.
7. **Brady, S., and J. Clardy.** 2000. Long-chain N-acetyl amino acid antibiotics isolated from heterologously expressed environmental DNA. J. Am. Chem. Soc. **122:**12903–12904.
8. **Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer.** 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. USA **99:**14250–14255.
9. **Cho, J. C., and J. M. Tiedje.** 2001. Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. Appl. Environ. Microbiol. **67:**3677–3682.
10. **Cho, J. C., and J. M. Tiedje.** 2002. Quantitative detection of microbial genes by using DNA microarrays. Appl. Environ. Microbiol. **68:**1425–1430.
11. **Cottrell, M. T., J. A. Moore, and D. L. Kirchman.** 1999. Chitinases from uncultured marine microorganisms. Appl. Environ. Microbiol. **65:**2553–2557.
12. **Eilers, H., J. Pernthaler, F. O. Glockner, and R. Amann.** 2000. Culturability and in situ abundance of pelagic bacteria from the North Sea. Appl. Environ. Microbiol. **66:**3044–3051.
13. **Ellis, R., I. Thompson, and M. Bailey.** 1995. Metabolic profiling as a means of characterizing plant-associated microbial communities. FEMS Microbiol. Ecol. **16:**9–18.
14. **Garcia, A. B., J. Engler, S. Iyer, T. Gerats, M. Van Montagu, and A. B. Caplan.** 1997. Effects of osmoprotectants upon NaCl stress in rice. Plant Physiol. **115:**159–169.
15. **Gillespie, D. E., S. F. Brady, A. D. Bettermann, N. P. Cianciotto, M. R. Liles, M. R. Rondon, J. Clardy, R. M. Goodman, and J. Handelsman.** 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl. Environ. Microbiol. **68:**4301–4306.
16. **Henne, A., R. Daniel, R. A. Schmitz, and G. Gottschalk.** 1999. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl. Environ. Microbiol. **65:**3901–3907.
17. **Henne, A., R. A. Schmitz, M. Bomeke, G. Gottschalk, and R. Daniel.** 2000. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. Appl. Environ. Microbiol. **66:** 3113–3116.
18. **Joyce, E. A., K. Chan, N. R. Salama, and S. Falkow.** 2002. Redefining bacterial populations: a post-genomic reformation. Nat. Rev. Genet. **3:**462–473.
19. **Kleihues, L., O. Lenz, M. Bernhard, T. Buhrke, and B. Friedrich.** 2000. The

H_2 sensor of *Ralstonia eutropha* is a member of the subclass of regulatory [NiFe] hydrogenases. J. Bacteriol. **182:**2716–2724.
20. **Lane, D.** 1991. 16S/23S rRNA sequencing, p. 115–175. *In* E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics. Wiley, New York, N.Y.
21. **Liles, M. R., B. F. Manske, S. B. Bintrim, J. Handelsman, and R. M. Goodman.** 2003. A census of rRNA genes and linked genomic sequences within a soil metagenomic library. Appl. Environ. Microbiol. **69:**2684–2691.
22. **Liu, Y. G., and R. F. Whittier.** 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics **25:**674–681.
23. **Lucito, R., J. West, A. Reiner, J. Alexander, D. Esposito, B. Mishra, S. Powers, L. Norton, and M. Wigler.** 2000. Detecting gene copy number fluctuations in tumor cells by microarray analysis of genomic representations. Genome Res. **10:**1726–1736.
24. **MacNeil, I. A., C. L. Tiong, C. Minor, P. R. August, T. H. Grossman, K. A. Loiacono, B. A. Lynch, T. Phillips, S. Narula, R. Sundaramoorthi, A. Tyler, T. Aldredge, H. Long, M. Gilman, D. Holt, and M. S. Osburne.** 2001. Expression and isolation of antimicrobial small molecules from soil DNA libraries. J. Mol. Microbiol. Biotechnol. **3:**301–308.
25. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
26. **Majernik, A., G. Gottschalk, and R. Daniel.** 2001. Screening of environmental DNA libraries for the presence of genes conferring Na^+ (Li^+)/H^+ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. J. Bacteriol. **183:**6645–6653.
27. **Makrigiorgos, G. M., S. Chakrabarti, Y. Zhang, M. Kaur, and B. D. Price.** 2002. A PCR-based amplification method retaining the quantitative difference between two complex genomes. Nat. Biotechnol. **20:**936–939.
28. **Mavrodi, D. V., O. V. Mavrodi, B. B. McSpadden-Gardener, B. B. Landa, D. M. Weller, and L. S. Thomashow.** 2002. Identification of differences in genome content among *phlD*-positive *Pseudomonas fluorescens* strains by using PCR-based subtractive hybridization. Appl. Environ. Microbiol. **68:** 5170–5176.
29. **Muyzer, G., T. Brinkhoff, U. Nübel, C. Santegoeds, H. Schäfer, and C. Wawer.** 1997. Denaturing gradient gel electrophoresis (DGGE) in microbial ecology, p. 1–27. *In* A. D. L. Akkermans, J. D. van Elsas, and F. J. de Bruijn (ed.), Molecular microbial ecology manual. Kluwer Academic Publishers, Dordrecht, The Netherlands.
30. **Pollack, J. R., C. M. Perou, A. A. Alizadeh, M. B. Eisen, A. Pergamenschikov, C. F. Williams, S. S. Jeffrey, D. Botstein, and P. O. Brown.** 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat. Genet. **23:**41–46.
31. **Reysenbach, A. L., L. J. Giver, G. S. Wickham, and N. R. Pace.** 1992. Differential amplification of rRNA genes by polymerase chain reaction. Appl. Environ. Microbiol. **58:**3417–3418.
32. **Romermann, D., J. Warrelmann, R. A. Bender, and B. Friedrich.** 1989. An *rpoN*-like gene of *Alcaligenes eutrophus* and *Pseudomonas facilis* controls expression of diverse metabolic pathways, including hydrogen oxidation. J. Bacteriol. **171:**1093–1099.
33. **Rondon, M. R., S. J. Raffel, R. M. Goodman, and J. Handelsman.** 1999. Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. Proc. Natl. Acad. Sci. USA **96:**6451–6455.
34. **Small, J., D. R. Call, F. J. Brockman, T. M. Straub, and D. P. Chandler.** 2001. Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. Appl. Environ. Microbiol. **67:**4708–4716.
35. **Stein, J. L., T. L. Marsh, K. Y. Wu, H. Shizuya, and E. F. DeLong.** 1996. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J. Bacteriol. **178:**591–599.
36. **Stoner, D. L., M. C. Geary, L. J. White, R. D. Lee, J. A. Brizzee, A. C. Rodman, and R. C. Rope.** 2001. Mapping microbial biodiversity. Appl. Environ. Microbiol. **67:**4324–4328.
37. **Suzuki, M. T., and S. J. Giovannoni.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. **62:**625–630.
38. **Urasaki, A., Y. Sekine, and E. Ohtsubo.** 2002. Transposition of cyanobacterium insertion element ISY*100* in *Escherichia coli*. J. Bacteriol. **184:**5104–5112.
39. **Venter, J. C., H. O. Smith, and L. Hood.** 1996. A new strategy for genome sequencing. Nature **381:**364–366.
40. **von Wintzingerode, F., U. B. Gobel, and E. Stackebrandt.** 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. **21:**213–229.
41. **Wang, G. C. and Y. Wang.** 1996. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiol. **142:**1107–1114.
42. **Wang, G. Y., E. Graziani, B. Waters, W. Pan, X. Li, J. McDermott, G. Meurer, G. Saxena, R. J. Andersen, and J. Davies.** 2000. Novel natural products from soil DNA libraries in a streptomycete host. Org. Lett. **2:**2401–2404.
43. **Wilson, K.** 1992. Preparation of genomic DNA from bacteria, p. 2–10–2–11. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Short protocols in molecular biology. John Wiley & Sons, New York, N.Y.
44. **Wilson, K. H., W. J. Wilson, J. L. Radosevich, T. Z. DeSantis, V. S. Viswanathan, T. A. Kuczmarski, and G. L. Andersen.** 2002. High-density microarray of small-subunit ribosomal DNA probes. Appl. Environ. Microbiol. **68:**2535–2541.
45. **Wu, L., D. K. Thompson, G. Li, R. A. Hurt, J. M. Tiedje, and J. Zhou.** 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl. Environ. Microbiol. **67:**5780–5790.